

REPRESENTATION OF SINGULAR INTEGRALS BY DYADIC OPERATORS, AND THE A_2 THEOREM

TUOMAS P. HYTÖNEN

Lecture notes of an intensive course at Universidad de Sevilla, Summer 2011.

ABSTRACT. These expository lectures present a self-contained proof of the A_2 theorem—the sharp weighted norm inequality for Calderón–Zygmund operators in $L^2(w)$ —, which is here formulated in such a way as to reveal some additional information implicit in the earlier papers. This added data gives at once a new weighted bound for powers of the Ahlfors–Beurling operator, discussed in the end. A key ingredient of the A_2 theorem is the probabilistic Dyadic Representation Theorem, for which a slightly simplified proof is given, avoiding conditional probabilities which were needed in the earlier arguments.

1. INTRODUCTION

The goal of the lectures is to prove the following A_2 theorem:

1.1. Theorem. *For any Calderón–Zygmund operator T on \mathbb{R}^d , any $w \in A_2$, and $f \in L^2(w)$, we have*

$$\|Tf\|_{L^2(w)} \leq C_T[w]_{A_2} \|f\|_{L^2(w)}.$$

The proof will proceed via the following steps, in the same order:

- Reduction to *dyadic shift operators*: every Calderón–Zygmund operator T has a representation in terms of these simpler operators, and hence it suffices to prove a similar claim for every dyadic shift S in place of T .
- Reduction to *testing conditions*: in order to have the full norm inequality

$$\|Sf\|_{L^2(w)} \leq C_S[w]_{A_2} \|f\|_{L^2(w)},$$

it suffices to have such an inequality for special test functions only:

$$\|S(1_Q w^{-1})\|_{L^2(w)} \leq C_S[w]_{A_2} \|1_Q w^{-1}\|_{L^2(w)},$$

$$\|S^*(1_Q w)\|_{L^2(w^{-1})} \leq C_S[w]_{A_2} \|1_Q w\|_{L^2(w^{-1})}.$$

- Verification of the testing conditions for S .

In the original proof of this theorem, in Summer 2010, the two reductions were done in a different order: the (quite complicated) reduction to testing condition was obtained for general Calderón–Zygmund operators by Pérez–Treil–Volberg [18]; my completion of the proof [6] then consisted of reducing these testing conditions for T to the testing conditions for S , and verifying the testing conditions for S as indicated in the last step above. The first two steps were assembled in the present order in our joint work [10], simplifying the overall argument: the reduction to dyadic operators and the verification of the testing conditions are essentially the

Date: August 26, 2011.

2010 Mathematics Subject Classification. 42B25, 42B35.

same, but the reduction to testing conditions is considerably simpler for the dyadic operators. The actual verification of the testing conditions, both in [6, 10] and in the present lectures, derives its main inspiration from the work of Lacey–Petermichl–Reguera [11].

Over the past year, different proofs and extensions of the A_2 theorem have appeared; see the final section for a discussion and selected references. Some proofs do not proceed via the testing conditions, but all known proofs so far do need the reduction to dyadic operators. Some ingredients from the newer proofs will be exploited to round up corners of the presentation here and there in these lectures.

2. PRELIMINARIES

The standard (or reference) system of dyadic cubes is

$$\mathcal{D}^0 := \{2^{-k}([0, 1]^d + m) : k \in \mathbb{Z}, m \in \mathbb{Z}^d\}.$$

We will need several dyadic systems, obtained by translating the reference system as follows. Let $\omega = (\omega_j)_{j \in \mathbb{Z}} \in (\{0, 1\}^d)^{\mathbb{Z}}$ and

$$I \dot{+} \omega := I + \sum_{j: 2^{-j} < \ell(I)} 2^{-j} \omega_j.$$

Then

$$\mathcal{D}^\omega := \{I \dot{+} \omega : I \in \mathcal{D}^0\},$$

and it is straightforward to check that \mathcal{D}^ω inherits the important nestedness property of \mathcal{D}^0 : if $I, J \in \mathcal{D}^\omega$, then $I \cap J \in \{I, J, \emptyset\}$. When the particular ω is unimportant, the notation \mathcal{D} is sometimes used for a generic dyadic system.

2.A. Haar functions. Any given dyadic system \mathcal{D} has a natural function system associated to it: the Haar functions. In one dimension, there are two Haar functions associated with an interval I : the non-cancellative $h_I^0 := |I|^{-1/2} 1_I$ and the cancellative $h_I^1 := |I|^{-1/2} (1_{I_\ell} - 1_{I_r})$, where I_ℓ and I_r are the left and right halves of I . In d dimensions, the Haar functions on a cube $I = I_1 \times \cdots \times I_d$ are formed of all the products of the one-dimensional Haar functions:

$$h_I^\eta(x) = h_{I_1 \times \cdots \times I_d}^{(\eta_1, \dots, \eta_d)}(x_1, \dots, x_d) := \prod_{i=1}^d h_{I_i}^{\eta_i}(x_i).$$

The non-cancellative $h_I^0 = |I|^{-1/2} 1_I$ has the same formula as in $d = 1$. All other $2^d - 1$ Haar functions h_I^η with $\eta \in \{0, 1\}^d \setminus \{0\}$ are cancellative, i.e., satisfy $\int h_I^\eta = 0$, since they are cancellative in at least one coordinate direction.

For a fixed \mathcal{D} , all the cancellative Haar functions h_I^η , $I \in \mathcal{D}$ and $\eta \in \{0, 1\}^d \setminus \{0\}$, form an orthonormal basis of $L^2(\mathbb{R}^d)$. Hence any function $f \in L^2(\mathbb{R}^d)$ has the orthogonal expansion

$$f = \sum_{I \in \mathcal{D}} \sum_{\eta \in \{0, 1\}^d \setminus \{0\}} \langle f, h_I^\eta \rangle h_I^\eta.$$

Since the different η 's seldom play any major role, this will be often abbreviated (with slight abuse of language) simply as

$$f = \sum_{I \in \mathcal{D}} \langle f, h_I \rangle h_I,$$

and the summation over η is understood implicitly.

2.B. Dyadic shifts. A dyadic shift with parameters $i, j \in \mathbb{N} := \{0, 1, 2, \dots\}$ is an operator of the form

$$Sf = \sum_{K \in \mathcal{D}} A_K f, \quad A_K f = \sum_{\substack{I, J \in \mathcal{D}: I, J \subseteq K \\ \ell(I) = 2^{-i} \ell(K) \\ \ell(J) = 2^{-j} \ell(K)}} a_{IJK} \langle f, h_I \rangle h_J,$$

where h_I is a Haar function on I (similarly h_J), and the a_{IJK} are coefficients with

$$|a_{IJK}| \leq \frac{\sqrt{|I||J|}}{|K|}.$$

It is also required that all subshifts

$$S_{\mathcal{Q}} = \sum_{K \in \mathcal{Q}} A_K, \quad \mathcal{Q} \subseteq \mathcal{D},$$

map $S_{\mathcal{Q}} : L^2(\mathbb{R}^d) \rightarrow L^2(\mathbb{R}^d)$ with norm at most one.

The shift is called cancellative, if all the h_I and h_J are cancellative; otherwise, it is called non-cancellative.

The notation A_K indicates an “averaging operator” on K . Indeed, from the normalization of the Haar functions, it follows that

$$|A_K f| \leq 1_K \int_K |f|$$

pointwise.

For cancellative shifts, the L^2 boundedness is automatic from the other conditions. This is a consequence of the following facts:

- The pointwise bound for each A_K implies that $\|A_K f\|_{L^p} \leq \|f\|_{L^p}$ for all $p \in [1, \infty]$; in particular, these components of S are uniformly bounded on L^2 with norm one. (This first point is true even in the non-cancellative case.)
- Let \mathbb{D}_K^i denote the orthogonal projection of L^2 onto $\text{span}\{h_I : I \subseteq K, \ell(I) = 2^{-i} \ell(K)\}$. When i is fixed, it follows readily that any two \mathbb{D}_K^i are orthogonal to each other. (This depends on the use of cancellative h_I .) Moreover, we have $A_K = \mathbb{D}_K^j A_K \mathbb{D}_K^i$. Then the boundedness of S follows from two applications of Pythagoras’ theorem with the uniform boundedness of the A_K in between.

A prime example of a non-cancellative shift (and the only one we need in these lectures) is the *dyadic paraproduct*

$$\Pi_b f = \sum_{K \in \mathcal{D}} \langle b, h_K \rangle \langle f \rangle_K h_K = \sum_{K \in \mathcal{D}} |K|^{-1/2} \langle b, h_K \rangle \cdot \langle f, h_K^0 \rangle h_K,$$

where $b \in \text{BMO}_d$ (the dyadic BMO space) and h_K is a cancellative Haar function. This is a dyadic shift with parameters $(i, j) = (0, 0)$, where $a_{IJK} = |K|^{-1/2} \langle b, h_K \rangle$ for $I = J = K$. The L^2 boundedness of the paraproduct, if and only if $b \in \text{BMO}_d$, is part of the classical theory. Actually, to ensure the normalization condition of the shift, it should be further required that $\|b\|_{\text{BMO}_d} \leq 1$.

2.C. Random dyadic systems; good and bad cubes. We obtain a notion of *random dyadic systems* by equipping the parameter set $\Omega := (\{0, 1\}^d)^\mathbb{Z}$ with the natural probability measure: each component ω_j has an equal probability 2^{-d} of taking any of the 2^d values in $\{0, 1\}^d$, and all components are independent of each other.

Let $\phi : [0, 1] \rightarrow [0, 1]$ be a fixed *modulus of continuity*: a strictly increasing function with $\phi(0) = 0$, $\phi(1) = 1$, and $t \mapsto \phi(t)/t$ decreasing (hence $\phi(t) \geq t$) with $\lim_{t \rightarrow 0} \phi(t)/t = \infty$. We further require the *Dini condition*

$$\int_0^1 \phi(t) \frac{dt}{t} < \infty.$$

Main examples include $\phi(t) = t^\gamma$ with $\gamma \in (0, 1)$ and

$$\phi(t) = \left(1 + \frac{1}{\gamma} \log \frac{1}{t}\right)^{-\gamma}, \quad \gamma > 1.$$

We also fix a (large) parameter $r \in \mathbb{Z}_+$. (How large, will be specified shortly.)

A cube $I \in \mathcal{D}^\omega$ is called bad if there exists $J \in \mathcal{D}^\omega$ such that $\ell(J) \geq 2^r \ell(I)$ and

$$\text{dist}(I, \partial J) \leq \phi\left(\frac{\ell(I)}{\ell(J)}\right) \ell(J) :$$

roughly, I is relatively close to the boundary of a much bigger cube.

2.1. Remark. This definition of good cubes goes back to Nazarov–Treil–Volberg [15] in the context of singular integrals with respect to non-doubling measures. They used the modulus of continuity $\phi(t) = t^\gamma$, where γ was chosen to depend on the dimension and the Hölder exponent of the Calderón–Zygmund kernel via

$$\gamma = \frac{\alpha}{2(d + \alpha)}.$$

This choice has become “canonical” in the subsequent literature, including the original proof of the A_2 theorem. However, other choices can also be made, as we do here.

We make some basic probabilistic observations related to badness. Let $I \in \mathcal{D}^0$ be a reference interval. The *position* of the translated interval

$$I \dot{+} \omega = I + \sum_{j: 2^{-j} < \ell(I)} 2^{-j} \omega_j,$$

by definition, depends only on ω_j for $2^{-j} < \ell(I)$. On the other hand, the *badness* of $I \dot{+} \omega$ depends on its *relative position* with respect to the bigger intervals

$$J \dot{+} \omega = J + \sum_{j: 2^{-j} < \ell(I)} 2^{-j} \omega_j + \sum_{j: \ell(I) \leq 2^{-j} < \ell(J)} 2^{-j} \omega_j.$$

The same translation component $\sum_{j: 2^{-j} < \ell(I)} 2^{-j} \omega_j$ appears in both $I \dot{+} \omega$ and $J \dot{+} \omega$, and so does not affect the relative position on these intervals. Thus this relative position, and hence the badness of I , depends only on ω_j for $2^{-j} \geq \ell(I)$. In particular:

2.2. Lemma. *For $I \in \mathcal{D}^0$, the position and badness of $I \dot{+} \omega$ are independent random variables.*

Another observation is the following: by symmetry and the fact that the condition of badness only involves relative position and size of different cubes, it readily follows that the probability of a particular cube $I + \omega$ being bad is equal for all cubes $I \in \mathcal{D}^0$:

$$\mathbb{P}_\omega(I + \omega \text{ bad}) = \pi_{\text{bad}} = \pi_{\text{bad}}(r, d, \phi).$$

The final observation concerns the value of this probability:

2.3. Lemma. *We have*

$$\pi_{\text{bad}} \leq 8d \int_0^{2^{-r}} \phi(t) \frac{dt}{t};$$

in particular, $\pi_{\text{bad}} < 1$ if $r = r(d, \phi)$ is chosen large enough.

With $r = r(d, \phi)$ chosen like this, we then have $\pi_{\text{good}} := 1 - \pi_{\text{bad}} > 0$, namely, good situations have positive probability!

Proof. Observe that in the definition of badness, we only need to consider those J with $I \subseteq J$. Namely, if I is close to the boundary of some bigger J , we can always find another dyadic J' of the same size as J which contains I , and then I will also be close to the boundary of J' . Hence we need to consider the relative position of I with respect to each $J \supset I$ with $\ell(J) = 2^k \ell(I)$ and $k = r, r+1, \dots$. For a fixed k , this relative position is determined by

$$\sum_{j: \ell(I) \leq 2^{-j} < 2^k \ell(I)} 2^{-j} \omega_j,$$

which has 2^{kd} different values with equal probability. These correspond to the subcubes of J of size $\ell(I)$.

Now bad position of I are those which are within distance $\phi(\ell(I)/\ell(J)) \cdot \ell(J)$ from the boundary. Since the possible position of the subcubes are discrete, being integer multiples of $\ell(I)$, the effective bad boundary region has depth

$$\begin{aligned} \left\lceil \phi\left(\frac{\ell(I)}{\ell(J)}\right) \frac{\ell(J)}{\ell(I)} \right\rceil \ell(I) &\leq \left(\phi\left(\frac{\ell(I)}{\ell(J)}\right) \frac{\ell(J)}{\ell(I)} + 1 \right) \ell(I) \\ &= \ell(J) \left(\phi\left(\frac{\ell(I)}{\ell(J)}\right) + \frac{\ell(I)}{\ell(J)} \right) \leq 2\ell(J) \phi\left(\frac{\ell(I)}{\ell(J)}\right), \end{aligned}$$

by using that $t \leq \phi(t)$.

The good region is the cube inside J , whose side-length is $\ell(J)$ minus twice the depth of the bad boundary region:

$$\ell(J) - 2 \left\lceil \phi\left(\frac{\ell(I)}{\ell(J)}\right) \frac{\ell(J)}{\ell(I)} \right\rceil \ell(I) \geq \ell(J) - 4\ell(J) \phi\left(\frac{\ell(I)}{\ell(J)}\right).$$

Hence the volume of the bad region is

$$\begin{aligned} |J| - \left(\ell(J) - 2 \left\lceil \phi\left(\frac{\ell(I)}{\ell(J)}\right) \frac{\ell(J)}{\ell(I)} \right\rceil \ell(I) \right)^d &\leq |J| \left(1 - \left(1 - 4\phi\left(\frac{\ell(I)}{\ell(J)}\right) \right)^d \right) \\ &\leq |J| \cdot 4d\phi\left(\frac{\ell(I)}{\ell(J)}\right) \end{aligned}$$

by the elementary inequality $(1 - \alpha)^d \geq 1 - \alpha d$ for $\alpha \in [0, 1]$. (We assume that r is at least so large that $4\phi(2^{-r}) \leq 1$.)

So the fraction of the bad region of the total volume is at most $4d\phi(\ell(I)/\ell(J)) = 4d\phi(2^{-k})$ for a fixed $k = r, r+1, \dots$. This gives the final estimate

$$\begin{aligned} \mathbb{P}_\omega(I \dot{+} \omega \text{ bad}) &\leq \sum_{k=r}^{\infty} 4d\phi(2^{-k}) = \sum_{k=r}^{\infty} 8d \frac{\phi(2^{-k})}{2^{-k}} 2^{-k-1} \\ &\leq \sum_{k=r}^{\infty} 8d \int_{2^{-k-1}}^{2^{-k}} \frac{\phi(t)}{t} dt = 8d \int_0^{2^{-r}} \phi(t) \frac{dt}{t}, \end{aligned}$$

where we used that $\phi(t)/t$ is decreasing in the last inequality. \square

3. THE DYADIC REPRESENTATION THEOREM

Let T be a Calderón–Zygmund operator on \mathbb{R}^d . That is, it acts on a suitable dense subspace of functions in $L^2(\mathbb{R}^d)$ (for the present purposes, this class should at least contain the indicators of cubes in \mathbb{R}^d) and has the kernel representation

$$Tf(x) = \int_{\mathbb{R}^d} K(x, y) f(y) dy, \quad x \notin \text{supp } f.$$

Moreover, the kernel should satisfy the *standard estimates*, which we here assume in a slightly more general form than usual, involving another modulus of continuity ψ , like the one considered above:

$$\begin{aligned} |K(x, y)| &\leq \frac{C_0}{|x - y|^d}, \\ |K(x, y) - K(x', y)| + |K(y, x) - K(y, x')| &\leq \frac{C_\psi}{|x - y|^d} \psi\left(\frac{|x - x'|}{|x - y|}\right) \end{aligned}$$

for all $x, x', y \in \mathbb{R}^d$ with $|x - y| > 2|x - x'|$. Let us denote the smallest admissible constants C_0 and C_ψ by $\|K\|_{CZ_0}$ and $\|K\|_{CZ_\psi}$. The classical standard estimates correspond to the choice $\psi(t) = t^\alpha$, $\alpha \in (0, 1]$, in which case we write $\|K\|_{CZ_\alpha}$ for $\|K\|_{CZ_\psi}$.

We say that T is a bounded Calderón–Zygmund operator, if in addition $T : L^2(\mathbb{R}^d) \rightarrow L^2(\mathbb{R}^d)$, and we denote its operator norm by $\|T\|_{L^2 \rightarrow L^2}$.

Let us agree that $|\cdot|$ stands for the ℓ^∞ norm on \mathbb{R}^d , i.e., $|x| := \max_{1 \leq i \leq d} |x_i|$. While the choice of the norm is not particularly important, this choice is slightly more convenient than the usual Euclidean norm when dealing with cubes as we will: e.g., the diameter of a cube in the ℓ^∞ norm is equal to its sidelength $\ell(Q)$.

Let us first formulate the dyadic representation theorem for general moduli of continuity, and then specialize it to the usual standard estimates. Define the following coefficients for $i, j \in \mathbb{N}$:

$$\tau(i, j) := \phi(2^{-\max\{i, j\}})^{-d} \psi(2^{-\max\{i, j\}} \phi(2^{-\max\{i, j\}})^{-1}),$$

if $\min\{i, j\} > 0$; and

$$\tau(i, j) := \Psi(2^{-\max\{i, j\}} \phi(2^{-\max\{i, j\}})^{-1}), \quad \Psi(t) := \int_0^t \psi(s) \frac{ds}{s},$$

if $\min\{i, j\} = 0$.

We assume that ϕ and ψ are such, that

$$(3.1) \quad \sum_{i, j=0}^{\infty} \tau(i, j) \approx \int_0^1 \frac{1}{\phi(t)^d} \psi\left(\frac{t}{\phi(t)}\right) \frac{dt}{t} + \int_0^1 \Psi\left(\frac{t}{\phi(t)}\right) \frac{dt}{t} < \infty.$$

This is the case, in particular, when $\psi(t) = t^\alpha$ (usual standard estimates) and $\phi(t) = (1 + a^{-1} \log t^{-1})^{-\gamma}$; then one checks that

$$\tau(i, j) \lesssim P(\max\{i, j\})2^{-\alpha \max\{i, j\}}, \quad P(j) = (1 + j)^{\gamma(d+\alpha)},$$

which clearly satisfies the required convergence. However, it is also possible to treat weaker forms of the standard estimates with a logarithmic modulus $\psi(t) = (1 + a^{-1} \log t^{-1})^{-\alpha}$. This might be of some interest for applications, but we do not pursue this line any further here.

3.2. Theorem. *Let T be a bounded Calderón–Zygmund operator with modulus of continuity satisfying the above assumption. Then it has an expansion, say for $f, g \in C_c^1(\mathbb{R}^d)$,*

$$\langle g, Tf \rangle = c \cdot (\|T\|_{L^2 \rightarrow L^2} + \|K\|_{CZ_\psi}) \cdot \mathbb{E}_\omega \sum_{i,j=0}^{\infty} \tau(i, j) \langle g, S_\omega^{ij} f \rangle,$$

where c is a dimensional constant and S_ω^{ij} is a dyadic shift of parameters (i, j) on the dyadic system \mathcal{D}^ω ; all of them except possibly S_ω^{00} are cancellative.

The first version of this theorem appeared in [6], and another one in [10]. The present proof is yet another variant of the same argument. It is slightly simpler in terms of the probabilistic tools that are used: no conditional probabilities are needed, although they were important for the original arguments.

In proving this theorem, we do not actually need to employ the full strength of the assumption that $T : L^2(\mathbb{R}^d) \rightarrow L^2(\mathbb{R}^d)$; rather it suffices to have the kernel conditions plus the following conditions of the $T1$ theorem of David–Journé:

$$|\langle 1_Q, T1_Q \rangle| \leq C_{WBP} |Q| \quad (\text{weak boundedness property}),$$

$$T1 \in \text{BMO}(\mathbb{R}^d), \quad T^*1 \in \text{BMO}(\mathbb{R}^d).$$

Let us denote the smallest C_{WBP} by $\|T\|_{WBP}$. Then we have the following more precise version of the representation:

3.3. Theorem. *Let T be a Calderón–Zygmund operator with modulus of continuity satisfying the above assumption. Then it has an expansion, say for $f, g \in C_c^1(\mathbb{R}^d)$,*

$$\begin{aligned} \langle g, Tf \rangle &= c \cdot (\|K\|_{CZ_0} + \|K\|_{CZ_\psi}) \mathbb{E}_\omega \sum_{\substack{i,j=0 \\ \max\{i,j\} > 0}}^{\infty} \tau(i, j) \langle g, S_\omega^{ij} f \rangle \\ &\quad + c \cdot (\|K\|_{CZ_0} + \|T\|_{WBP}) \mathbb{E}_\omega \langle g, S_\omega^{00} f \rangle + \mathbb{E}_\omega \langle g, \Pi_{T1}^\omega f \rangle + \mathbb{E}_\omega \langle g, (\Pi_{T^*1}^\omega)^* f \rangle \end{aligned}$$

where S_ω^{ij} is a cancellative dyadic shift of parameters (i, j) on the dyadic system \mathcal{D}^ω , and Π_b^ω is a dyadic paraproduct on the dyadic system \mathcal{D}^ω associated with the BMO-function $b \in \{T1, T^*1\}$.

3.4. Remark. Note that $\Pi_b^\omega = \|b\|_{\text{BMO}} \cdot S_b^\omega$, where $S_b^\omega = \Pi_b^\omega / \|b\|_{\text{BMO}}$ is a shift with the correct normalization. Hence, writing everything in terms of normalized shifts, as in Theorem 3.2, we get the factor $\|T1\|_{\text{BMO}} \lesssim \|T\|_{L^2 \rightarrow L^2} + \|K\|_{CZ_\psi}$ in the second-to-last term, and $\|T^*1\|_{\text{BMO}} \lesssim \|T\|_{L^2 \rightarrow L^2} + \|K\|_{CZ_\psi}$ in the last one. The proof will also show that both occurrences of the factor $\|K\|_{CZ_0}$ could be replaced by $\|T\|_{L^2 \rightarrow L^2}$, giving the statement of Theorem 3.2 (since trivially $\|T\|_{WBP} \leq \|T\|_{L^2 \rightarrow L^2}$).

As a by-product, Theorem 3.3 delivers a proof of the $T1$ theorem: under the above assumptions, the operator T is already bounded on $L^2(\mathbb{R}^d)$. Namely, all the dyadic shifts S_ω^{ij} are uniformly bounded on $L^2(\mathbb{R}^d)$ by definition, and the convergence condition (3.1) ensures that so is their average representing the operator T . This by-product proof of the $T1$ theorem is not a coincidence, since the proof of Theorems 3.2 and 3.3 was actually inspired by the proof of the $T1$ theorem for non-doubling measures due to Nazarov–Treil–Volberg [15] and its vector-valued extension [5].

A key to the proof of the dyadic representation is a random expansion of T in terms of Haar functions h_I , where the bad cubes are avoided:

3.5. Proposition.

$$\langle g, Tf \rangle = \frac{1}{\pi_{\text{good}}} \mathbb{E}_\omega \sum_{I, J \in \mathcal{D}^\omega} 1_{\text{good}}(\text{smaller}\{I, J\}) \cdot \langle g, h_J \rangle \langle h_J, Th_I \rangle \langle h_I, f \rangle,$$

where

$$\text{smaller}\{I, J\} := \begin{cases} I & \text{if } \ell(I) \leq \ell(J), \\ J & \text{if } \ell(J) > \ell(I). \end{cases}$$

Proof. Recall that

$$f = \sum_{I \in \mathcal{D}^0} \langle f, h_{I+\omega} \rangle h_{I+\omega}$$

for any fixed $\omega \in \Omega$; and we can also take the expectation \mathbb{E}_ω of both sides of this identity.

Let

$$1_{\text{good}}(I+\omega) := \begin{cases} 1, & \text{if } I+\omega \text{ is good,} \\ 0, & \text{else} \end{cases}$$

We make use of the above random Haar expansion of f , multiply and divide by

$$\pi_{\text{good}} = \mathbb{P}_\omega(I+\omega \text{ good}) = \mathbb{E}_\omega 1_{\text{good}}(I+\omega),$$

and use the independence from Lemma 2.2 to get:

$$\begin{aligned} \langle g, Tf \rangle &= \mathbb{E}_\omega \sum_I \langle g, Th_{I+\omega} \rangle \langle h_{I+\omega}, f \rangle \\ &= \frac{1}{\pi_{\text{good}}} \sum_I \mathbb{E}_\omega [1_{\text{good}}(I+\omega)] \mathbb{E}_\omega [\langle g, Th_{I+\omega} \rangle \langle h_{I+\omega}, f \rangle] \\ &= \frac{1}{\pi_{\text{good}}} \mathbb{E}_\omega \sum_I 1_{\text{good}}(I+\omega) \langle g, Th_{I+\omega} \rangle \langle h_{I+\omega}, f \rangle \\ &= \frac{1}{\pi_{\text{good}}} \mathbb{E}_\omega \sum_{I, J} 1_{\text{good}}(I+\omega) \langle g, h_{J+\omega} \rangle \langle h_{J+\omega}, Th_{I+\omega} \rangle \langle h_{I+\omega}, f \rangle. \end{aligned}$$

On the other hand, using independence again in half of this double sum, we have

$$\begin{aligned}
& \frac{1}{\pi_{\text{good}}} \sum_{\ell(I) > \ell(J)} \mathbb{E}_\omega [1_{\text{good}}(I \dot{+} \omega) \langle g, h_{J \dot{+} \omega} \rangle \langle h_{J \dot{+} \omega}, Th_{I \dot{+} \omega} \rangle \langle h_{I \dot{+} \omega}, f \rangle] \\
&= \frac{1}{\pi_{\text{good}}} \sum_{\ell(I) > \ell(J)} \mathbb{E}_\omega [1_{\text{good}}(I \dot{+} \omega)] \mathbb{E}_\omega [\langle g, h_{J \dot{+} \omega} \rangle \langle h_{J \dot{+} \omega}, Th_{I \dot{+} \omega} \rangle \langle h_{I \dot{+} \omega}, f \rangle] \\
&= \mathbb{E}_\omega \sum_{\ell(I) > \ell(J)} \langle g, h_{J \dot{+} \omega} \rangle \langle h_{J \dot{+} \omega}, Th_{I \dot{+} \omega} \rangle \langle h_{I \dot{+} \omega}, f \rangle,
\end{aligned}$$

and hence

$$\begin{aligned}
\langle g, Tf \rangle &= \frac{1}{\pi_{\text{good}}} \mathbb{E}_\omega \sum_{\ell(I) \leq \ell(J)} 1_{\text{good}}(I \dot{+} \omega) \langle g, h_{J \dot{+} \omega} \rangle \langle h_{J \dot{+} \omega}, Th_{I \dot{+} \omega} \rangle \langle h_{I \dot{+} \omega}, f \rangle \\
&\quad + \mathbb{E}_\omega \sum_{\ell(I) > \ell(J)} \langle g, h_{J \dot{+} \omega} \rangle \langle h_{J \dot{+} \omega}, Th_{I \dot{+} \omega} \rangle \langle h_{I \dot{+} \omega}, f \rangle.
\end{aligned}$$

Comparison with the basic identity

$$(3.6) \quad \langle g, Tf \rangle = \mathbb{E}_\omega \sum_{I, J} \langle g, h_{J \dot{+} \omega} \rangle \langle h_{J \dot{+} \omega}, Th_{I \dot{+} \omega} \rangle \langle h_{I \dot{+} \omega}, f \rangle$$

shows that

$$\begin{aligned}
& \mathbb{E}_\omega \sum_{\ell(I) \leq \ell(J)} \langle g, h_{J \dot{+} \omega} \rangle \langle h_{J \dot{+} \omega}, Th_{I \dot{+} \omega} \rangle \langle h_{I \dot{+} \omega}, f \rangle \\
&= \frac{1}{\pi_{\text{good}}} \mathbb{E}_\omega \sum_{\ell(I) \leq \ell(J)} 1_{\text{good}}(I \dot{+} \omega) \langle g, h_{J \dot{+} \omega} \rangle \langle h_{J \dot{+} \omega}, Th_{I \dot{+} \omega} \rangle \langle h_{I \dot{+} \omega}, f \rangle.
\end{aligned}$$

Symmetrically, we also have

$$\begin{aligned}
& \mathbb{E}_\omega \sum_{\ell(I) > \ell(J)} \langle g, h_{J \dot{+} \omega} \rangle \langle h_{J \dot{+} \omega}, Th_{I \dot{+} \omega} \rangle \langle h_{I \dot{+} \omega}, f \rangle \\
&= \frac{1}{\pi_{\text{good}}} \mathbb{E}_\omega \sum_{\ell(I) > \ell(J)} 1_{\text{good}}(J \dot{+} \omega) \langle g, h_{J \dot{+} \omega} \rangle \langle h_{J \dot{+} \omega}, Th_{I \dot{+} \omega} \rangle \langle h_{I \dot{+} \omega}, f \rangle,
\end{aligned}$$

and this completes the proof. \square

This is essentially the end of probability in this proof. Henceforth, we can simply concentrate on the summation inside \mathbb{E}_ω , for a fixed value of $\omega \in \Omega$, and manipulate it into the required form. Moreover, we will concentrate on the half of the sum with $\ell(J) \geq \ell(I)$, the other half being handled symmetrically. We further divide this sum into the following parts:

$$\begin{aligned}
\sum_{\ell(I) \leq \ell(J)} &= \sum_{\text{dist}(I, J) > \ell(J)\phi(\ell(I)/\ell(J))} + \sum_{I \subsetneq J} + \sum_{I=J} + \sum_{\substack{\text{dist}(I, J) \leq \ell(J)\phi(\ell(I)/\ell(J)) \\ I \cap J = \emptyset}} \\
&=: \sigma_{\text{out}} + \sigma_{\text{in}} + \sigma_{=} + \sigma_{\text{near}}.
\end{aligned}$$

In order to recognize these series as sums of dyadic shifts, we need to locate, for each pair (I, J) appearing here, a common dyadic ancestor which contains both of them. The existence of such containing cubes, with control on their size, is provided by the following:

3.7. Lemma. *If $I \in \mathcal{D}$ is good and $J \in \mathcal{D}$ is a disjoint ($J \cap I = \emptyset$) cube with $\ell(J) \geq \ell(I)$, then there exists $K \supseteq I \cup J$ which satisfies*

$$\begin{aligned} \ell(K) &\leq 2^r \ell(I), & \text{if} & \quad \text{dist}(I, J) \leq \ell(J) \phi\left(\frac{\ell(I)}{\ell(J)}\right), \\ \ell(K) \phi\left(\frac{\ell(I)}{\ell(K)}\right) &\leq 2^r \text{dist}(I, J), & \text{if} & \quad \text{dist}(I, J) > \ell(J) \phi\left(\frac{\ell(I)}{\ell(J)}\right). \end{aligned}$$

Proof. Let us start with the following initial observation: if $K \in \mathcal{D}$ satisfies $I \subseteq K$, $J \subset K^c$, and $\ell(K) \geq 2^r \ell(I)$, then

$$\ell(K) \phi\left(\frac{\ell(I)}{\ell(K)}\right) < \text{dist}(I, \partial K) = \text{dist}(I, K^c) \leq \text{dist}(I, J).$$

Case $\text{dist}(I, J) \leq \ell(J) \phi(\ell(I)/\ell(J))$. As $I \cap J = \emptyset$, we have $\text{dist}(I, J) = \text{dist}(I, \partial J)$, and since I is good, this implies $\ell(J) < 2^r \ell(I)$. Let $K = I^{(r)}$, and assume for contradiction that $J \subset K^c$. Then the initial observation implies that

$$\ell(K) \phi\left(\frac{\ell(I)}{\ell(K)}\right) < \text{dist}(I, J) \leq \ell(J) \phi\left(\frac{\ell(I)}{\ell(J)}\right).$$

Dividing both sides by $\ell(I)$ and recalling that $\phi(t)/t$ is decreasing, this implies that $\ell(K) < \ell(J)$, a contradiction with $\ell(K) = 2^r \ell(I) > \ell(J)$. Hence $J \not\subset K^c$, and since $\ell(J) < \ell(K)$, this implies that $J \subset K$.

Case $\text{dist}(I, J) > \ell(J) \phi(\ell(I)/\ell(J))$. Consider the minimal $K \supset I$ with $\ell(K) \geq 2^r \ell(I)$ and $\text{dist}(I, J) \leq \ell(K) \phi(\ell(I)/\ell(K))$. (Since $\phi(t)/t \rightarrow \infty$ as $t \rightarrow 0$, this bound holds for all large enough K .) Then (since $\phi(t)/t$ is decreasing) $\ell(K) > \ell(J)$, and by the initial observation, $J \not\subset K^c$. Hence either $J \subset K$, and it suffices to estimate $\ell(K)$.

By the minimality of K , there holds at least one of

$$\frac{1}{2} \ell(K) < 2^r \ell(I) \quad \text{or} \quad \frac{1}{2} \ell(K) \phi\left(\frac{\ell(I)}{\frac{1}{2} \ell(K)}\right) < \text{dist}(I, J),$$

and the latter immediately implies that $\ell(K) \phi(\ell(I)/\ell(K)) < 2 \text{dist}(I, J)$. In the first case, since $\ell(I) \leq \ell(J) \leq \ell(K)$, we have

$$\ell(K) \phi\left(\frac{\ell(I)}{\ell(K)}\right) \leq 2^r \ell(I) \phi\left(\frac{\ell(I)}{\ell(K)}\right) \leq 2^r \ell(J) \phi\left(\frac{\ell(I)}{\ell(J)}\right) < 2^r \text{dist}(I, J),$$

so the required bound is true in each case. \square

We denote the minimal such K by $I \vee J$, thus

$$I \vee J := \bigcap_{K \supseteq I \cup J} K.$$

3.A. Separated cubes, σ_{out} . We reorganize the sum σ_{out} with respect to the new summation variable $K = I \vee J$, as well as the relative size of I and J with respect to K :

$$\sigma_{\text{out}} = \sum_{j=1}^{\infty} \sum_{i=j}^{\infty} \sum_K \sum_{\substack{\text{dist}(I, J) > \ell(J) \phi(\ell(I)/\ell(J)) \\ I \vee J = K \\ \ell(I) = 2^{-i} \ell(K), \ell(J) = 2^{-j} \ell(K)}}.$$

Note that we can start the summation from 1 instead of 0, since the disjointness of I and J implies that $K = I \vee J$ must be strictly larger than either of I and

J . The goal is to identify the quantity in parentheses as a decaying factor times a cancellative averaging operator with parameters (i, j) .

3.8. Lemma. *For I and J appearing in σ_{out} , we have*

$$|\langle h_J, Th_I \rangle| \lesssim \|K\|_{CZ_\psi} \frac{\sqrt{|I||J|}}{|K|} \phi\left(\frac{\ell(I)}{\ell(K)}\right)^{-d} \psi\left(\frac{\ell(I)}{\ell(K)}\right) \phi\left(\frac{\ell(I)}{\ell(K)}\right)^{-1}, \quad K = I \vee J.$$

Proof. Using the cancellation of h_I , standard estimates, and Lemma 3.7

$$\begin{aligned} |\langle h_J, Th_I \rangle| &= \left| \iint h_J(x) K(x, y) h_I(y) \, dy \, dx \right| \\ &= \left| \iint h_J(x) [K(x, y) - K(x, y_I)] h_I(y) \, dy \, dx \right| \\ &\lesssim \|K\|_{CZ_\psi} \iint |h_J(x)| \frac{1}{\text{dist}(I, J)^d} \psi\left(\frac{\ell(I)}{\text{dist}(I, J)}\right) |h_I(y)| \, dy \, dx \\ &= \|K\|_{CZ_\psi} \frac{1}{\text{dist}(I, J)^d} \psi\left(\frac{\ell(I)}{\text{dist}(I, J)}\right) \|h_J\|_1 \|h_I\|_1 \\ &\lesssim \|K\|_{CZ_\psi} \frac{1}{\ell(K)^d} \phi\left(\frac{\ell(I)}{\ell(K)}\right)^{-d} \psi\left(\frac{\ell(I)}{\ell(K)}\right) \phi\left(\frac{\ell(I)}{\ell(K)}\right)^{-1} \sqrt{|J|} \sqrt{|I|}. \quad \square \end{aligned}$$

3.9. Lemma.

$$\begin{aligned} &\sum_{\substack{\text{dist}(I, J) > \ell(J) \phi(\ell(I)/\ell(J)) \\ I \vee J = K \\ \ell(I) = 2^{-i} \ell(K) \leq \ell(J) = 2^{-j} \ell(K)}} 1_{\text{good}}(I) \cdot \langle g, h_J \rangle \langle h_J, Th_I \rangle \langle h_I, f \rangle \\ &= \|K\|_{CZ_\psi} \phi(2^{-i})^{-d} \psi(2^{-i} \phi(2^{-i})^{-1}) \langle g, A_K^{ij} f \rangle, \end{aligned}$$

where A_K^{ij} is a cancellative averaging operator with parameters (i, j) .

Proof. By the previous lemma, substituting $\ell(I)/\ell(K) = 2^{-i}$,

$$|\langle h_J, Th_I \rangle| \lesssim \|K\|_{CZ_\psi} \frac{\sqrt{|I||J|}}{|K|} \phi(2^{-i})^{-d} \psi(2^{-i} \phi(2^{-i})^{-1}),$$

and the first factor is precisely the required size of the coefficients of A_K^{ij} . \square

Summarizing, we have

$$\sigma_{\text{out}} = \|K\|_{CZ_\psi} \sum_{j=1}^{\infty} \sum_{i=j}^{\infty} \phi(2^{-i})^{-d} \psi(2^{-i} \phi(2^{-i})^{-1}) \langle g, S^{ij} f \rangle.$$

3.B. Contained cubes, σ_{in} . When $I \subsetneq J$, then I is contained in some subcube of J , which we denote by J_I .

$$\begin{aligned} \langle h_J, Th_I \rangle &= \langle 1_{J_I^c} h_J, Th_I \rangle + \langle 1_{J_I} h_J, Th_I \rangle \\ &= \langle 1_{J_I^c} h_J, Th_I \rangle + \langle h_J \rangle_{J_I} \langle 1_{J_I}, Th_I \rangle \\ &= \langle 1_{J_I^c} (h_J - \langle h_J \rangle_{J_I}), Th_I \rangle + \langle h_J \rangle_{J_I} \langle 1, Th_I \rangle, \end{aligned}$$

where we noticed that h_J is constant on $J_I \supseteq I$.

3.10. Lemma.

$$|\langle 1_{J_I^c}(h_J - \langle h_J \rangle_{J_I}), Th_I \rangle| \lesssim (\|K\|_{CZ_0} + \|K\|_{CZ_\psi}) \left(\frac{|I|}{|J|} \right)^{1/2} \Psi \left(\frac{\ell(I)}{\ell(J)} \phi \left(\frac{\ell(I)}{\ell(J)} \right)^{-1} \right),$$

where

$$\Psi(r) := \int_0^r \psi(t) \frac{dt}{t},$$

and $\|K\|_{CZ_0}$ could be alternatively replaced by $\|T\|_{L^2 \rightarrow L^2}$.

Proof.

$$|\langle 1_{J_I^c}(h_J - \langle h_J \rangle_{J_I}), Th_I \rangle| \leq 2\|h_J\|_\infty \int_{J_I^c} |Th_I(x)| dx,$$

where $\|h_J\|_\infty = |J|^{-1/2}$.

Case $\ell(I) \geq 2^{-r}\ell(J)$. We have

$$\begin{aligned} \int_{J_I^c} |Th_I(x)| dx &\leq \int_{3I \setminus I} \left| \int K(x, y) h_I(y) dy \right| dx \\ &\quad + \int_{(3I)^c} \left| \int [K(x, y) - K(x, y_I)] h_I(y) dy \right| dx \\ &\lesssim \|K\|_{CZ_0} \int_{3I \setminus I} \int_I \frac{1}{|x - y|^d} dy dx \|h_I\|_\infty \\ &\quad + \|K\|_{CZ_\psi} \int_{(3I)^c} \frac{1}{\text{dist}(x, I)^d} \psi \left(\frac{\ell(I)}{\text{dist}(x, I)} \right) \|h_I\|_1 dx \\ &\lesssim \|K\|_{CZ_0} |I| \|h_I\|_\infty + \|K\|_{CZ_\psi} \int_{\ell(I)}^\infty \frac{1}{r^d} \psi \left(\frac{\ell(I)}{r} \right) r^{d-1} dr \|h_I\|_1 \\ &= \|K\|_{CZ_0} |I|^{1/2} + \|K\|_{CZ_\psi} \int_0^1 \psi(t) \frac{dt}{t} |I|^{1/2} \\ &\lesssim (\|K\|_{CZ_0} + \|K\|_{CZ_\psi}) |I|^{1/2} \end{aligned}$$

by the Dini condition for ψ in the last step.

Alternatively, the part giving the factor $\|K\|_{CZ_0}$ could have been estimated by

$$\int_{3I \setminus I} \left| \int K(x, y) h_I(y) dy \right| dx \leq |3I \setminus I|^{1/2} \|Th_I\|_2 \lesssim |I|^{1/2} \|T\|_{L^2 \rightarrow L^2}.$$

Case $\ell(I) < 2^{-r}\ell(J)$. Since $I \subseteq J_I$ is good, we have

$$\text{dist}(I, J_I^c) > \ell(J_I) \phi \left(\frac{\ell(I)}{\ell(J_I)} \right) \gtrsim \ell(J) \phi \left(\frac{\ell(I)}{\ell(J)} \right)$$

and hence

$$\begin{aligned} \int_{J_I^c} |Th_I(x)| dx &\lesssim \|K\|_{CZ_\psi} \int_{J_I^c} \frac{1}{d(x, I)^d} \psi \left(\frac{\ell(I)}{\text{dist}(x, I)} \right) \|h_I\|_1 dx \\ &\lesssim \|K\|_{CZ_\psi} \int_{\ell(J) \phi(\ell(I)/\ell(J))}^\infty \frac{1}{r^d} \psi \left(\frac{\ell(I)}{r} \right) r^{d-1} dr \cdot \|h_I\|_1 \\ &= \|K\|_{CZ_\psi} \int_0^{\ell(I)/\ell(J) \cdot \phi(\ell(I)/\ell(J))^{-1}} \psi(t) \frac{dt}{t} \cdot |I|^{1/2}. \quad \square \end{aligned}$$

Now we can organize

$$\sigma'_{\text{in}} := \sum_J \sum_{I \subsetneq J} \langle g, h_J \rangle \langle 1_{J_I^c} (h_J - \langle h_J \rangle_{J_I}), Th_I \rangle \langle h_I, f \rangle = \sum_{i=1}^{\infty} \sum_J \sum_{\substack{I \subset J \\ \ell(I)=2^{-i}\ell(J)}},$$

and the inner sum is recognized as

$$(\|K\|_{CZ_0} + \|K\|_{CZ_\psi}) \Psi(2^{-i} \phi(2^{-i})^{-1}) \langle g, A_J^{i0} f \rangle,$$

or with $\|T\|_{L^2 \rightarrow L^2}$ in place of $\|K\|_{CZ_0}$, for a cancellative averaging operator of type $(i, 0)$.

On the other hand,

$$\begin{aligned} \sigma''_{\text{in}} &:= \sum_J \sum_{I \subsetneq J} \langle g, h_J \rangle \langle h_J \rangle_I \langle 1, Th_I \rangle \langle h_I, f \rangle \\ &= \sum_I \left\langle \sum_{J \supsetneq I} \langle g, h_J \rangle h_J \right\rangle_I \langle 1, Th_I \rangle \langle h_I, f \rangle \\ &= \sum_I \langle g \rangle_I \langle T^* 1, h_I \rangle \langle h_I, f \rangle \\ &= \left\langle \sum_I \langle g \rangle_I \langle T^* 1, h_I \rangle h_I, f \right\rangle =: \langle \Pi_{T^*1} g, f \rangle = \langle g, \Pi_{T^*1}^* f \rangle. \end{aligned}$$

Here Π_{T^*1} is the *paraproduct*, a non-cancellative shift composed of the non-cancellative averaging operators

$$A_I g = \langle T^* 1, h_I \rangle \langle g \rangle_I h_I = |I|^{-1/2} \langle T^* 1, h_I \rangle \cdot \langle g, h_I^0 \rangle h_I$$

of type $(0, 0)$.

Summarizing, we have

$$\begin{aligned} \sigma_{\text{in}} &= \sigma'_{\text{in}} + \sigma''_{\text{in}} \\ &= (\|K\|_{CZ_0} + \|K\|_{CZ_\psi}) \sum_{i=1}^{\infty} \Psi(2^{-i} \phi(2^{-i})^{-1}) \langle g, S^{i0} f \rangle + \langle \Pi_{T^*1} g, f \rangle, \end{aligned}$$

where $\Psi(t) = \int_0^t \psi(s) ds/s$, and $\|K\|_{CZ_0}$ could be replaced by $\|T\|_{L^2 \rightarrow L^2}$. Note that if we wanted to write Π_{T^*1} in terms of a shift with correct normalization, we should divide and multiply it by $\|T^* 1\|_{\text{BMO}}$, thus getting a shift times the factor $\|T^* 1\|_{\text{BMO}} \lesssim \|T\|_{L^2} + \|K\|_{CZ_\psi}$.

3.C. Near-by cubes, $\sigma_{=}$ and σ_{near} . We are left with the sums $\sigma_{=}$ of equal cubes $I = J$, as well as σ_{near} of disjoint near-by cubes with $\text{dist}(I, J) \leq \ell(J) \phi(\ell(I)/\ell(J))$. Since I is good, this necessarily implies that $\ell(I) > 2^{-r} \ell(J)$. Then, for a given J , there are only boundedly many related I in this sum.

3.11. Lemma.

$$|\langle h_J, Th_I \rangle| \lesssim \|K\|_{CZ_0} + \delta_{IJ} \|T\|_{WBP}.$$

Note that if we used the L^2 -boundedness of T instead of the CZ_0 and WBP conditions (as is done in Theorem 3.2, we could also estimate simply

$$|\langle h_J, Th_I \rangle| \leq \|h_J\|_2 \|T\|_{L^2 \rightarrow L^2} \|h_I\|_2 = \|T\|_{L^2 \rightarrow L^2}.$$

Proof. For disjoint cubes, we estimate directly

$$\begin{aligned} |\langle h_J, Th_I \rangle| &\lesssim \|K\|_{CZ_0} \int_J \int_I \frac{1}{|x-y|^d} dy dx \|h_J\|_\infty \|h_I\|_\infty \\ &\leq \|K\|_{CZ_0} \int_J \int_{3J \setminus J} \frac{1}{|x-y|^d} dy dx |J|^{-1/2} |I|^{-1/2} \\ &\lesssim \|K\|_{CZ_0} |J| |J|^{-1/2} |J|^{-1/2} = \|K\|_{CZ_0}, \end{aligned}$$

since $|I| \approx |J|$.

For $J = I$, let I_i be its dyadic children. Then

$$\begin{aligned} |\langle h_I, Th_I \rangle| &\leq \sum_{i,j=1}^{2^d} |\langle h_I \rangle_{I_i} \langle h_I \rangle_{I_j} \langle 1_{I_i}, T 1_{I_j} \rangle| \\ &\lesssim \|K\|_{CZ_0} \sum_{j \neq i} |I|^{-1} \int_{I_i} \int_{I_j} \frac{1}{|x-y|^d} dx dy + \sum_i |I|^{-1} |\langle 1_{I_i}, T 1_{I_i} \rangle| \\ &\lesssim \|K\|_{CZ_0} + \|T\|_{WBP}, \end{aligned}$$

by the same estimate as earlier for the first term, and the weak boundedness property for the second. \square

With this lemma, the sum σ_+ is recognized as a cancellative dyadic shift of type $(0, 0)$ as such:

$$\begin{aligned} \sigma_+ &= \sum_{I \in \mathcal{D}} 1_{\text{good}}(I) \cdot \langle g, h_I \rangle \langle h_I, Th_I \rangle \langle h_I, f \rangle \\ &= (\|K\|_{CZ_0} + \|T\|_{WBP}) \langle g, S^{00} f \rangle, \end{aligned}$$

where the factor in front could also be replaced by $\|T\|_{L^2 \rightarrow L^2}$.

For I and J participating in σ_{near} , we conclude from Lemma 3.7 that $K := I \vee J$ satisfies $\ell(K) \leq 2^r \ell(I)$, and hence we may organize

$$\sigma_{\text{near}} = \sum_{i=1}^r \sum_{j=1}^i \sum_K \sum_{\substack{I, J: I \vee J = K \\ \text{dist}(I, J) \leq \ell(J) \phi(\ell(I)/\ell(J)) \\ \ell(I) = 2^{-i} \ell(K) \\ \ell(J) = 2^{-j} \ell(K)}},$$

and the innermost sum is recognized as $\|K\|_{CZ_0} \langle g, A_K^{ij} f \rangle$ for some cancellative averaging operator of type (i, j) .

Summarizing, we have

$$\sigma_+ + \sigma_{\text{near}} = (\|K\|_{CZ_0} + \|T\|_{WBP}) \langle g, S^{00} f \rangle + \|K\|_{CZ_0} \sum_{j=1}^r \sum_{i=j}^r \langle g, S^{ij} f \rangle,$$

where S^{00} and S^{ij} are cancellative dyadic shifts, and the factor $(\|K\|_{CZ_0} + \|T\|_{WBP})$ could also be replaced by $\|T\|_{L^2 \rightarrow L^2}$.

3.D. Synthesis. We have checked that

$$\begin{aligned} & \sum_{\ell(I) \leq \ell(J)} 1_{\text{good}}(I) \langle g, h_J \rangle \langle h_J, Th_I \rangle \langle h_I, f \rangle \\ &= (\|K\|_{CZ_0} + \|K\|_{CZ_\psi}) \left(\sum_{1 \leq j \leq i < \infty} \phi(2^{-i})^{-d} \psi(2^{-i} \phi(2^{-i})^{-1}) \langle g, S^{ij} f \rangle \right. \\ &\quad \left. + \sum_{1 \leq i < \infty} \Psi(2^{-i} \phi(2^{-i})^{-1}) \langle g, S^{i0} f \rangle \right) \\ &\quad + (\|K\|_{CZ_0} + \|T\|_{WBP}) \langle g, S^{00} f \rangle + \langle g, \Pi_{T^*1}^* f \rangle \end{aligned}$$

where $\Psi(t) = \int_0^t \psi(s) ds/s$, Π_{T^*1} is a paraproduct—a non-cancellative shift of type $(0, 0)_-$, and all other S^{ij} is a cancellative dyadic shifts of type (i, j) .

By symmetry (just observing that the cubes of equal size contributed precisely to the presence of the cancellative shifts of type (i, i) , and that the dual of a shift of type (i, j) is a shift of type (j, i)), it follows that

$$\begin{aligned} & \sum_{\ell(I) > \ell(J)} 1_{\text{good}}(J) \langle g, h_J \rangle \langle h_J, Th_I \rangle \langle h_I, f \rangle \\ &= (\|K\|_{CZ_0} + \|K\|_{CZ_\psi}) \left(\sum_{1 \leq i < j < \infty} \phi(2^{-j})^{-d} \psi(2^{-j} \phi(2^{-j})^{-1}) \langle g, S^{ij} f \rangle \right. \\ &\quad \left. + \sum_{1 \leq j < \infty} \Psi(2^{-j} \phi(2^{-j})^{-1}) \langle g, S^{0j} f \rangle \right) + \langle g, \Pi_{T1} f \rangle \end{aligned}$$

so that altogether

$$\begin{aligned} & \sum_{I, J} 1_{\text{good}}(\min\{I, J\}) \langle g, h_J \rangle \langle h_J, Th_I \rangle \langle h_I, f \rangle \\ &= (\|K\|_{CZ_0} + \|K\|_{CZ_\psi}) \left(\sum_{i=1}^{\infty} \Psi(2^{-i} \phi(2^{-i})^{-1}) (\langle g, S^{i0} f \rangle + \langle g, S^{0i} f \rangle) \right. \\ &\quad \left. + \sum_{i, j=1}^{\infty} \phi(2^{-\max(i, j)})^{-d} \psi(2^{-\max(i, j)} \phi(2^{-\max(i, j)})^{-1}) \langle g, S^{ij} f \rangle \right) \\ &\quad + (\|K\|_{CZ_0} + \|T\|_{WBP}) \langle g, S^{00} f \rangle + \langle g, \Pi_{T1} f \rangle + \langle g, \Pi_{T^*1}^* f \rangle, \end{aligned}$$

and this completes the proof of Theorem 3.2.

4. TWO-WEIGHT THEORY FOR DYADIC SHIFTS

Before proceeding further, it is convenient to introduce a useful trick due to E. Sawyer. Let σ be an everywhere positive, finitely-valued function. Then $f \in L^p(w)$ if and only if $\phi = f/\sigma \in L^p(\sigma^p w)$, and they have equal norms in the respective spaces. Hence an inequality

$$(4.1) \quad \|Tf\|_{L^p(w)} \leq N \|f\|_{L^p(w)} \quad \forall f \in L^p(w)$$

is equivalent to

$$\|T(\phi\sigma)\|_{L^p(w)} \leq N \|\phi\sigma\|_{L^p(w)} = N \|\phi\|_{L^p(\sigma^p w)} \quad \forall \phi \in L^p(\sigma^p w).$$

This is true for any σ , and we now choose it in such a way that $\sigma^p w = \sigma$, i.e., $\sigma = w^{-1/(p-1)} = w^{1-p'}$, where p' is the dual exponent. So finally (4.1) is equivalent

to

$$\|T(\phi\sigma)\|_{L^p(w)} \leq N\|\phi\|_{L^p(\sigma)} \quad \forall \phi \in L^p(\sigma).$$

This formulation has the advantage that the norm on the right and the operator

$$T(\phi\sigma)(x) = \int K(x, y)\phi(y) \cdot \sigma(y) dy$$

involve integration with respect to the same measure σ . In particular, the A_2 theorem is equivalent to

$$\|T(f\sigma)\|_{L^2(w)} \leq c_T[w]_{A_2}\|f\|_{L^2(\sigma)}$$

for all $f \in L^2(w)$, for all $w \in A_2$ and $\sigma = w^{-1}$. But once we know this, we can also study this two-weight inequality on its own right, for two general measures w and σ , which need not be related by the pointwise relation $\sigma(x) = 1/w(x)$.

4.2. Theorem. *Let σ and w be two locally finite measures with*

$$[w, \sigma]_{A_2} := \sup_Q \frac{w(Q)\sigma(Q)}{|Q|^2} < \infty.$$

Then a dyadic shift S of type (i, j) satisfies $S(\sigma) : L^2(\sigma) \rightarrow L^2(w)$ if and only if

$$\mathfrak{S} := \sup_Q \frac{\|1_Q S(\sigma 1_Q)\|_{L^2(w)}}{\sigma(Q)^{1/2}}, \quad \mathfrak{S}^* := \sup_Q \frac{\|1_Q S^*(w 1_Q)\|_{L^2(\sigma)}}{w(Q)^{1/2}}$$

are finite, and in this case

$$\|S(\sigma)\|_{L^2(\sigma) \rightarrow L^2(w)} \lesssim (1 + \kappa)(\mathfrak{S} + \mathfrak{S}^*) + (1 + \kappa)^2[w, \sigma]_{A_2}^{1/2},$$

where $\kappa = \max\{i, j\}$.

This result from my work with Pérez, Treil, and Volberg [10] was preceded by an analogous qualitative version due to Nazarov, Treil, and Volberg [16].

The proof depends on decomposing functions in the spaces $L^2(w)$ and $L^2(\sigma)$ in terms of expansions similar to the Haar expansion in $L^2(\mathbb{R}^d)$. Let \mathbb{D}_I^σ be the orthogonal projection of $L^2(\sigma)$ onto its subspace of functions supported on I , constant on the subcubes of I , and with vanishing integral with respect to $d\sigma$. Then any two \mathbb{D}_I^σ are orthogonal to each other. Under the additional assumption that the σ measure of quadrants of \mathbb{R}^d is infinite, we have the expansion

$$f = \sum_{Q \in \mathcal{Q}} \mathbb{D}_Q^\sigma f$$

for all $f \in L^2(\sigma)$, and Pythagoras' theorem says that

$$\|f\|_{L^2(\sigma)} = \left(\sum_{Q \in \mathcal{Q}} \|\mathbb{D}_Q^\sigma f\|_{L^2(\sigma)}^2 \right)^{1/2}.$$

(These formulae needs a slight adjustment if the σ measure of quadrants is finite; Theorem 4.2 remains true without this extra assumption.) Let us also write

$$\mathbb{D}_K^{\sigma, i} := \sum_{\substack{I \subseteq K \\ \ell(I) = 2^{-i}\ell(K)}} \mathbb{D}_I^\sigma.$$

For a fixed $i \in \mathbb{N}$, these are also orthogonal to each other, and the above formulae generalize to

$$f = \sum_{Q \in \mathcal{D}} \mathbb{D}_Q^{\sigma, i} f, \quad \|f\|_{L^2(\sigma)} = \left(\sum_{Q \in \mathcal{D}} \|\mathbb{D}_Q^{\sigma, i} f\|_{L^2(\sigma)}^2 \right)^{1/2}.$$

The proof is in fact very similar in spirit to that of Theorem 3.2; it is another $T1$ argument, but now with respect to the measures σ and w in place of the Lebesgue measure. We hence expand

$$\langle g, S(\sigma f) \rangle_w = \sum_{Q, R \in \mathcal{D}} \langle \mathbb{D}_R^w g, S(\sigma \mathbb{D}_Q^\sigma f) \rangle_w, \quad f \in L^2(\sigma), \quad g \in L^2(w),$$

and estimate the matrix coefficients

$$\begin{aligned} \langle \mathbb{D}_R^w g, S(\sigma \mathbb{D}_Q^\sigma f) \rangle_w &= \sum_K \langle \mathbb{D}_R^w g, A_K(\sigma \mathbb{D}_Q^\sigma f) \rangle_w \\ (4.3) \quad &= \sum_K \sum_{I, J \subseteq K} a_{IJK} \langle \mathbb{D}_R^w g, h_J \rangle_w \langle h_I, \mathbb{D}_Q^\sigma f \rangle_\sigma. \end{aligned}$$

For $\langle h_I, \mathbb{D}_Q^\sigma f \rangle_\sigma \neq 0$, there must hold $I \cap Q \neq \emptyset$, thus $I \subseteq Q$ or $Q \subsetneq I$. But in the latter case h_I is constant on Q , while $\int \mathbb{D}_Q^\sigma f \cdot \sigma = 0$, so the pairing vanishes even in this case. Thus the only nonzero contributions come from $I \subseteq Q$, and similarly from $J \subseteq R$. Since $I, J \subseteq K$, there holds

$$(I \subseteq Q \subsetneq K \quad \text{or} \quad K \subseteq Q) \quad \text{and} \quad (J \subseteq R \subsetneq K \quad \text{or} \quad K \subseteq R).$$

4.A. Disjoint cubes. Suppose now that $Q \cap R = \emptyset$, and let K be among those cubes for which A_K gives a nontrivial contribution above. Then it cannot be that $K \subseteq Q$, since this would imply that $Q \cap R \supseteq K \cap J = J \neq \emptyset$, and similarly it cannot be that $K \subseteq R$. Thus $Q, R \subsetneq K$, and hence

$$Q \vee R \subseteq K.$$

Then

$$\begin{aligned} |\langle \mathbb{D}_R^w g, S(\sigma \mathbb{D}_Q^\sigma f) \rangle_w| &\leq \sum_{K \supseteq Q \vee R} |\langle \mathbb{D}_R^w g, A_K(\sigma \mathbb{D}_Q^\sigma f) \rangle_w| \\ &\lesssim \sum_{K \supseteq Q \vee R} \frac{\|\mathbb{D}_R^w g\|_{L^1(w)} \|\mathbb{D}_Q^\sigma f\|_{L^1(\sigma)}}{|K|} \\ &\lesssim \frac{\|\mathbb{D}_R^w g\|_{L^1(w)} \|\mathbb{D}_Q^\sigma f\|_{L^1(\sigma)}}{|Q \vee R|} \end{aligned}$$

On the other hand, we have $Q \supseteq I$, $R \supseteq J$ for some $I, J \subseteq K$ with $\ell(I) = 2^{-i}\ell(K)$ and $\ell(J) = 2^{-j}\ell(K)$. Hence $2^{-i}\ell(K) \leq \ell(Q)$ and $2^{-j}\ell(K) \leq \ell(R)$, and thus

$$Q \vee R \subseteq K \subseteq Q^{(i)} \cap R^{(j)}.$$

Now it is possible to estimate the total contribution of the part of the matrix with $Q \cap R = \emptyset$. Let $P := Q \vee R$ be a new auxiliary summation variable. Then $Q, R \subset P$, and $\ell(Q) = 2^{-a}\ell(P)$, $\ell(R) = 2^{-b}\ell(P)$ where $a = 1, \dots, i$, $b = 1, \dots, j$. Thus

$$\sum_{\substack{Q, R \in \mathcal{D} \\ Q \cap R = \emptyset}} |\langle \mathbb{D}_R^w g, S(\sigma \mathbb{D}_Q^\sigma f) \rangle_w|$$

$$\begin{aligned}
&\lesssim \sum_{a=1}^i \sum_{b=1}^j \sum_{P \in \mathcal{D}} \frac{1}{|P|} \sum_{\substack{Q, R \in \mathcal{D}: Q \vee R = P \\ \ell(Q) = 2^{-a} \ell(P) \\ \ell(R) = 2^{-b} \ell(P)}} \|\mathbb{D}_R^w g\|_{L^1(\sigma)} \|\mathbb{D}_Q^\sigma f\|_{L^1(w)} \\
&\leq \sum_{a,b=1}^{i,j} \sum_{P \in \mathcal{D}} \frac{1}{|P|} \sum_{\substack{R \subseteq P \\ \ell(R) = 2^{-b} \ell(P)}} \|\mathbb{D}_R^w g\|_{L^1(\sigma)} \sum_{\substack{Q \subseteq P \\ \ell(Q) = 2^{-a} \ell(P)}} \|\mathbb{D}_Q^\sigma f\|_{L^1(\sigma)} \\
&= \sum_{a,b=1}^{i,j} \sum_{P \in \mathcal{D}} \frac{1}{|P|} \left\| \sum_{\substack{R \subseteq P \\ \ell(R) = 2^{-b} \ell(P)}} \mathbb{D}_R^w g \right\|_{L^1(\sigma)} \left\| \sum_{\substack{Q \subseteq P \\ \ell(Q) = 2^{-a} \ell(P)}} \mathbb{D}_Q^\sigma f \right\|_{L^1(\sigma)} \\
&\quad \text{(by disjoint supports)} \\
&= \sum_{a,b=1}^{i,j} \sum_{P \in \mathcal{D}} \frac{1}{|P|} \|\mathbb{D}_P^{w,j} g\|_{L^1(w)} \|\mathbb{D}_P^{\sigma,i} f\|_{L^1(\sigma)} \\
&\leq \sum_{a,b=1}^{i,j} \sum_{P \in \mathcal{D}} \frac{\sigma(P)^{1/2} w(P)^{1/2}}{|P|} \|\mathbb{D}_P^{w,j} g\|_{L^2(w)} \|\mathbb{D}_P^{\sigma,i} f\|_{L^2(\sigma)} \\
&\leq \sum_{a,b=1}^{i,j} [w, \sigma]_{A_2}^{1/2} \left(\sum_{P \in \mathcal{D}} \|\mathbb{D}_P^{w,j} g\|_{L^2(w)}^2 \right)^{1/2} \left(\sum_{P \in \mathcal{D}} \|\mathbb{D}_P^{\sigma,i} f\|_{L^2(\sigma)}^2 \right)^{1/2} \\
&\leq ij [w, \sigma]_{A_2}^{1/2} \|g\|_{L^2(w)} \|f\|_{L^2(\sigma)}.
\end{aligned}$$

4.B. Deeply contained cubes. Consider now the part of the sum with $Q \subset R$ and $\ell(Q) < 2^{-i} \ell(R)$. (The part with $R \subset Q$ and $\ell(R) < 2^{-j} \ell(Q)$ would be handled in a symmetrical manner.)

4.4. Lemma. *For all $Q \subset R$ with $\ell(Q) < 2^{-i} \ell(R)$, we have*

$$\langle \mathbb{D}_R^w g, S(\sigma \mathbb{D}_Q^\sigma f) \rangle_w = \langle \mathbb{D}_R^w g \rangle_{Q^{(i)}} \langle S^*(w 1_{Q^{(i)}}), \mathbb{D}_Q^\sigma f \rangle_\sigma,$$

where further

$$\mathbb{D}_Q^\sigma S^*(w 1_{Q^{(i)}}) = \mathbb{D}_Q^\sigma S^*(w 1_P) \quad \text{for any } P \supseteq Q^{(i)}.$$

Recall that $\mathbb{D}_Q^\sigma = (\mathbb{D}_Q^\sigma)^2 = (\mathbb{D}_Q^\sigma)^*$ is an orthogonal projection on $L^2(\sigma)$, so that it can be moved to either or both sides of the pairing $\langle \cdot, \cdot \rangle_\sigma$.

Proof. Recall formula (4.3). If $\langle h_I, \mathbb{D}_Q^\sigma f \rangle_\sigma$ is nonzero, then $I \subseteq Q$, and hence

$$J \subseteq K = I^{(i)} \subseteq Q^{(i)} \subsetneq R$$

for all J participating in the same A_K as I . Thus $\mathbb{D}_R^w g$ is constant on $Q^{(i)}$, hence

$$\begin{aligned}
\langle \mathbb{D}_R^w g, A_K(\sigma \mathbb{D}_Q^\sigma f) \rangle_w &= \langle 1_{Q^{(i)}} \mathbb{D}_R^w g, A_K(\sigma \mathbb{D}_Q^\sigma f) \rangle_w \\
&= \langle \mathbb{D}_R^w g \rangle_{Q^{(i)}}^w \langle 1_{Q^{(i)}}, A_K(\sigma \mathbb{D}_Q^\sigma f) \rangle_w \\
&= \langle \mathbb{D}_R^w g \rangle_{Q^{(i)}}^w \langle A_K^*(w 1_{Q^{(i)}}), \mathbb{D}_Q^\sigma f \rangle_\sigma.
\end{aligned}$$

Moreover, for any $P \supseteq Q^{(i)} \supseteq K$,

$$\begin{aligned} \langle \mathbb{D}_Q^\sigma A_K^*(w1_{Q^{(i)}}), f \rangle_\sigma &= \langle 1_{Q^{(i)}}, A_K(\sigma \mathbb{D}_Q^\sigma f) \rangle_w \\ &= \int A_K(\sigma \mathbb{D}_Q^\sigma f) w \\ &= \langle 1_P, A_K(\sigma \mathbb{D}_Q^\sigma f) \rangle_w = \langle \mathbb{D}_Q^\sigma A_K^*(w1_P), f \rangle_\sigma. \end{aligned}$$

Summing these equalities over all relevant K , and using $S = \sum_K A_K$, gives the claim. \square

By the lemma, we can then manipulate

$$\begin{aligned} \sum_{\substack{Q, R: Q \subseteq R \\ \ell(Q) < 2^{-i} \ell(R)}} \langle \mathbb{D}_R^w g, S(\sigma \mathbb{D}_Q^\sigma f) \rangle_w &= \sum_Q \left(\sum_{R \supsetneq Q^{(i)}} \langle \mathbb{D}_R^w g \rangle_{Q^{(i)}}^w \right) \langle S^*(w1_{Q^{(i)}}), \mathbb{D}_Q^\sigma f \rangle_\sigma \\ &= \sum_Q \langle g \rangle_{Q^{(i)}}^w \langle S^*(w1_{Q^{(i)}}), \mathbb{D}_Q^\sigma f \rangle_\sigma \\ &= \sum_R \langle g \rangle_R^w \left\langle S^*(w1_R), \sum_{\substack{Q \subseteq R \\ \ell(Q) = 2^{-i} \ell(R)}} \mathbb{D}_Q^\sigma f \right\rangle_\sigma \\ &= \sum_R \langle g \rangle_R^w \left\langle S^*(w1_R), \mathbb{D}_R^{\sigma, i} f \right\rangle_\sigma, \end{aligned}$$

where $\langle g \rangle_R^w := w(R)^{-1} \int_R g \cdot w$ is the average of g on R with respect to the w measure.

By using the properties of the pairwise orthogonal projections $\mathbb{D}_R^{\sigma, i}$ on $L^2(\sigma)$, the above series may be estimated as follows:

$$\begin{aligned} &\left| \sum_{\substack{Q, R: Q \subseteq R \\ \ell(Q) < 2^{-i} \ell(R)}} \langle \mathbb{D}_R^w g, S(\sigma \mathbb{D}_Q^\sigma f) \rangle_w \right| \\ &\leq \sum_R |\langle g \rangle_R^w| \|\mathbb{D}_R^{\sigma, i} S^*(w1_R)\|_{L^2(\sigma)} \|\mathbb{D}_R^{\sigma, i} f\|_{L^2(\sigma)} \\ &\leq \left(\sum_R |\langle g \rangle_R^w|^2 \|\mathbb{D}_R^{\sigma, i} S^*(w1_R)\|_{L^2(\sigma)}^2 \right)^{1/2} \left(\sum_R \|\mathbb{D}_R^{\sigma, i} f\|_{L^2(\sigma)}^2 \right)^{1/2}, \end{aligned}$$

where the last factor is equal to $\|f\|_{L^2(w)}$.

The first factor on the right is handled by the dyadic Carleson embedding theorem: It follows from the second equality of Lemma 4.4, namely $\mathbb{D}_Q^\sigma S^*(w1_Q^{(i)}) = \mathbb{D}_Q^\sigma S^*(w1_P)$ for all $P \supseteq Q^{(i)}$, that $\mathbb{D}_R^{\sigma, i} S^*(w1_R) = \mathbb{D}_Q^\sigma S^*(w1_P)$ for all $P \subseteq R$. Hence, we have

$$\begin{aligned} \sum_{R \subseteq P} \|\mathbb{D}_R^{\sigma, i} S^*(w1_R)\|_{L^2(\sigma)}^2 &= \sum_{R \subseteq P} \|\mathbb{D}_R^{\sigma, i} (1_P S^*(w1_P))\|_{L^2(\sigma)}^2 \\ &\leq \|1_P S^*(w1_P)\|_{L^2(\sigma)}^2 \lesssim \mathfrak{S}_*^2 \sigma(P) \end{aligned}$$

by the (dual) testing estimate for the dyadic shifts. By the Carleson embedding theorem, it then follows that

$$\left(\sum_R |\langle g \rangle_R^w|^2 \|\mathbb{D}_R^{\sigma,i} S^*(w1_R)\|_{L^2(\sigma)}^2 \right)^{1/2} \lesssim \mathfrak{S}_* \|g\|_{L^2(\sigma)},$$

and the estimation of the deeply contained cubes is finished.

4.C. Contained cubes of comparable size. It remains to estimate

$$\sum_{\substack{Q, R: Q \subseteq R \\ \ell(Q) \geq 2^{-i} \ell(R)}} \langle \mathbb{D}_R^w g, S(\sigma \mathbb{D}_Q^\sigma f) \rangle_w;$$

the sum over $R \subsetneq Q$ with $\ell(R) \geq 2^{-j} \ell(Q)$ would be handled in a symmetric manner. The sum of interest may be written as

$$\sum_{a=0}^i \sum_R \sum_{\substack{Q \subseteq R \\ \ell(Q) = 2^{-a} \ell(R)}} \langle \mathbb{D}_R^w g, S(\sigma \mathbb{D}_Q^\sigma f) \rangle_w = \sum_{a=0}^i \sum_R \langle \mathbb{D}_R^w g, S(\sigma \mathbb{D}_R^{\sigma,i} f) \rangle_w,$$

and

$$\langle \mathbb{D}_R^w g, S(\sigma \mathbb{D}_R^{\sigma,i} f) \rangle_w = \sum_{k=1}^{2^d} \langle \mathbb{D}_R^w g \rangle_{R_k} \langle S^*(w1_{R_k}), \mathbb{D}_R^{\sigma,i} f \rangle_\sigma$$

where the R_k are the 2^d dyadic children of R , and $\langle \mathbb{D}_R^w g \rangle_{R_k}$ is the constant valued of $\mathbb{D}_R^w g$ on R_k . Now

$$\langle S^*(w1_{R_k}), \mathbb{D}_R^{\sigma,i} f \rangle_\sigma = \langle 1_{R_k} S^*(w1_{R_k}), \mathbb{D}_R^{\sigma,i} f \rangle_\sigma + \langle S^*(w1_{R_k}), 1_{R_k^c} \mathbb{D}_R^{\sigma,i} f \rangle_\sigma,$$

where

$$|\langle 1_{R_k} S^*(w1_{R_k}), \mathbb{D}_R^{\sigma,i} f \rangle_\sigma| \leq \mathfrak{S}_* w(R_k)^{1/2} \|\mathbb{D}_R^{\sigma,i} f\|_{L^2(\sigma)}$$

and, observing that only those A_K^* where K intersects both R_k and R_k^c contribute to the second part,

$$\begin{aligned} |\langle S^*(w1_{R_k}), 1_{R_k^c} \mathbb{D}_R^{\sigma,i} f \rangle_\sigma| &= \left| \sum_{K \supsetneq R_k} \langle A_K^*(w1_{R_k}), 1_{R_k^c} \mathbb{D}_R^{\sigma,i} f \rangle_\sigma \right| \\ &\lesssim \sum_{K \supsetneq R_k} \frac{1}{|K|} w(R_k) \|\mathbb{D}_R^{\sigma,i} f\|_{L^1(\sigma)} \\ &\lesssim \frac{1}{|R|} w(R_k) \sigma(R)^{1/2} \|\mathbb{D}_R^{\sigma,i} f\|_{L^1(\sigma)} \\ &\leq \frac{w(R)^{1/2} \sigma(R)^{1/2}}{|R|} w(R_k)^{1/2} \|\mathbb{D}_R^{\sigma,i} f\|_{L^2(\sigma)} \\ &\leq [w, \sigma]_{A_2} w(R_k)^{1/2} \|\mathbb{D}_R^{\sigma,i} f\|_{L^2(\sigma)}. \end{aligned}$$

It follows that

$$|\langle S^*(w1_{R_k}), \mathbb{D}_R^{\sigma,i} f \rangle_\sigma| \lesssim (\mathfrak{S}_* + [w, \sigma]_{A_2}) w(R_k)^{1/2} \|\mathbb{D}_R^{\sigma,i} f\|_{L^2(\sigma)}$$

and hence

$$|\langle \mathbb{D}_R^w g, S(\sigma \mathbb{D}_R^{\sigma,i} f) \rangle_w| \lesssim (\mathfrak{S}_* + [w, \sigma]_{A_2}) \|\mathbb{D}_R^w g\|_{L^2(w)} \|\mathbb{D}_R^{\sigma,i} f\|_{L^2(\sigma)}$$

Finally,

$$\begin{aligned}
& \sum_{a=0}^i \sum_R |\langle \mathbb{D}_R^w g, S(\sigma \mathbb{D}_R^{\sigma, i} f) \rangle_w| \\
& \lesssim (\mathfrak{S}_* + [w, \sigma]_{A_2}) \sum_{a=0}^i \left(\sum_R \|\mathbb{D}_R^w g\|_{L^2(\sigma)}^2 \right)^{1/2} \left(\sum_R \|\mathbb{D}_R^{\sigma, i} f\|_{L^2(\sigma)}^2 \right)^{1/2} \\
& \leq (1+i)(\mathfrak{S}_* + [w, \sigma]_{A_2}) \|g\|_{L^2(w)} \|f\|_{L^2(\sigma)}.
\end{aligned}$$

The symmetric case with $R \subset Q$ with $\ell(R) \geq 2^{-j}\ell(Q)$ similarly yields the factor $(1+j)(\mathfrak{S} + [w, \sigma]_{A_2})$. This completes the proof of Theorem 4.2.

5. FINAL DECOMPOSITIONS: VERIFICATION OF THE TESTING CONDITIONS

We now turn to the estimation of the testing constant

$$\mathfrak{S} := \sup_{Q \in \mathcal{D}} \frac{\|1_Q S(\sigma 1_Q)\|_{L^2(w)}}{\sigma(Q)^{1/2}}.$$

Bounding \mathfrak{S}_* is analogous by exchanging the roles of w and σ .

5.A. Several splittings. First observe that

$$1_Q S(\sigma 1_Q) = 1_Q \sum_{K: K \cap Q \neq \emptyset} A_K(\sigma 1_Q) = 1_Q \sum_{K \subseteq Q} A_K(\sigma 1_Q) + 1_Q \sum_{K \not\subseteq Q} A_K(\sigma 1_Q).$$

The second part is immediate to estimate even pointwise by

$$|1_Q A_K(\sigma 1_Q)| \leq 1_Q \frac{\sigma(Q)}{|K|}, \quad \sum_{K \not\subseteq Q} \frac{1}{|K|} \leq \frac{1}{|Q|},$$

and hence its $L^2(w)$ norm is bounded by

$$\left\| 1_Q \frac{\sigma(Q)}{|Q|} \right\|_{L^2(w)} = \frac{w(Q)^{1/2} \sigma(Q)}{|Q|} \leq [w, \sigma]_{A_2} \sigma(Q)^{1/2}.$$

So it remains to concentrate on $K \subseteq Q$, and we perform several consecutive splittings of this collection of cubes. First, we **separate scales** by introducing the splitting according to the $\kappa + 1$ possible values of $\log_2 \ell(K) \bmod (\kappa + 1)$. We denote a generic choice of such a collection by

$$\mathcal{K} = \mathcal{K}_k := \{K \subseteq Q : \log_2 \ell(K) \equiv k \bmod (\kappa + 1)\},$$

where k is arbitrary but fixed. (We will drop the subscript k , since its value plays no role in the subsequent argument.) Next, we **freeze the A_2 characteristic** by setting

$$\mathcal{K}^a := \left\{ K \in \mathcal{K} : 2^{a-1} < \frac{w(K)\sigma(K)}{|K|} \leq 2^a \right\}, \quad a \in \mathbb{Z}, \quad a \leq \lceil \log_2 [w, \sigma]_{A_2} \rceil,$$

where $\lceil \cdot \rceil$ means rounding up to the next integer.

In the next step, we **choose the principal cubes** $P \in \mathcal{P}^a \subseteq \mathcal{K}^a$. Let \mathcal{P}_0^a consist of all maximal cubes in \mathcal{K}^a , and inductively \mathcal{P}_{p+1}^a consist of all maximal $P' \in \mathcal{K}^a$ such that

$$P' \subset P \in \mathcal{P}_p^a, \quad \frac{\sigma(P')}{|P'|} > 2 \frac{\sigma(P)}{|P|}.$$

Finally, let $\mathcal{P}^a := \bigcup_{p=0}^{\infty} \mathcal{P}_p^a$. For each $K \in \mathcal{K}^a$, let $\Pi^a(K)$ denote the minimal $P \in \mathcal{P}^a$ such that $K \subseteq P$. Then we set

$$\mathcal{K}^a(P) := \{K \in \mathcal{K}^a : \Pi^a(K) = P\}, \quad P \in \mathcal{P}^a.$$

Note that $\sigma(K)/|K| \leq 2\sigma(P)/|P|$ for all $K \in \mathcal{K}^a(P)$, which allows us to **freeze the σ -to-Lebesgue measure ratio** by the final subcollections

$$\mathcal{K}_b^a(P) := \left\{K \in \mathcal{K}^a(P) : 2^{-b} < \frac{\sigma(K)}{|K|} \frac{|P|}{\sigma(P)} \leq 2^{1-b}\right\}, \quad b \in \mathbb{N}.$$

We have

$$\begin{aligned} \{K \in \mathcal{D} : K \subseteq Q\} &= \bigcup_{k=0}^{\kappa} \mathcal{K}_k, \quad \mathcal{K}_k = \mathcal{K} = \bigcup_{a \leq \lceil \log_2[w, \sigma]_{A_2} \rceil} \mathcal{K}^a, \\ \mathcal{K}^a &= \bigcup_{P \in \mathcal{P}^a} \mathcal{K}^a(P), \quad \mathcal{K}^a(P) = \bigcup_{b=0}^{\infty} \mathcal{K}_b^a(P), \end{aligned}$$

where all unions are disjoint. Note that we drop the reference to the separation-of-scales parameter k , since this plays no role in the forthcoming arguments. Recalling the notation for subshifts $S_{\mathcal{Q}} = \sum_{K \in \mathcal{Q}} A_K$, this splitting of collections of cubes leads to the splitting of the function

$$\sum_{K \subseteq Q} A_K(\sigma 1_Q) = \sum_{k=0}^{\kappa} \sum_{a \leq \lceil \log_2[w, \sigma]_{A_2} \rceil} \sum_{P \in \mathcal{P}^a} \sum_{b=0}^{\infty} S_{\mathcal{K}_b^a(P)}(\sigma 1_Q).$$

On the level of the function, we split one more time to write

$$\begin{aligned} S_{\mathcal{K}_b^a(P)}(\sigma 1_Q) &= \sum_{n=0}^{\infty} 1_{E_b^a(P, n)} S_{\mathcal{K}_b^a(P)}(\sigma 1_Q), \\ E_b^a(P, n) &:= \{x \in \mathbb{R}^d : n2^{-b} \langle \sigma \rangle_P < |S_{\mathcal{K}_b^a(P)}(\sigma 1_Q)(x)| \leq (n+1)2^{-b} \langle \sigma \rangle_P\}. \end{aligned}$$

This final splitting, from [8], is not strictly ‘necessary’ in that it was not part of the original argument in [6], nor its predecessor in [11], which made instead more careful use of the cubes where $S_{\mathcal{K}_b^a(P)}(\sigma 1_Q)$ stays constant; however, it now seems that this splitting provides another simplification of the argument.

Now all relevant cancellation is inside the functions $S_{\mathcal{K}_b^a(P)}(\sigma 1_Q)$, so that we can simply estimate by the triangle inequality:

$$\begin{aligned} &\left| \sum_{K \subseteq Q} A_K(\sigma 1_Q) \right| \\ &\leq \sum_{k=0}^{\kappa} \sum_a \sum_{P \in \mathcal{P}^a} \sum_{b=0}^{\infty} \sum_{n=0}^{\infty} (1+n) 2^{-b} \langle \sigma \rangle_P 1_{\{|S_{\mathcal{K}_b^a(P)}(\sigma 1_Q)| > n2^{-b} \langle \sigma \rangle_P\}}, \end{aligned}$$

and

$$\begin{aligned} &\left\| \sum_{K \subseteq Q} A_K(\sigma 1_Q) \right\|_{L^2(w)} \\ &\leq \sum_{k=0}^{\kappa} \sum_a \sum_{b=0}^{\infty} 2^{-b} \sum_{n=0}^{\infty} (1+n) \left\| \sum_{P \in \mathcal{P}^a} \langle \sigma \rangle_P 1_{\{|S_{\mathcal{K}_b^a(P)}(\sigma 1_Q)| > n2^{-b} \langle \sigma \rangle_P\}} \right\|_{L^2(w)}. \end{aligned}$$

Obviously, we will need good estimates to be able to sum up these infinite series.

Write the last norm as

$$\left(\int \left[\sum_{P \in \mathcal{P}^a} \langle \sigma \rangle_P 1_{\{|S_{\mathcal{K}_b^a(P)}(\sigma 1_Q)| > n2^{-b} \langle \sigma \rangle_P\}}(x) \right]^2 dw(x) \right)^{1/2},$$

observe that

$$\{|S_{\mathcal{K}_b^a(P)}(\sigma 1_Q)| > n2^{-b} \langle \sigma \rangle_P\} \subseteq P,$$

and look at the integrand at a fixed point $x \in \mathbb{R}^d$. At this point we sum over a subset of those values of $\langle \sigma \rangle_P$ where the principal cube $P \ni x$. Let P_0 be the smallest cube such that $|S_{\mathcal{K}_b^a(P)}| > n2^{-b} \langle \sigma \rangle_P$, let P_1 be the next smallest, and so on. Then $\langle \sigma \rangle_{P_m} < 2^{-1} \langle \sigma \rangle_{P_{m-1}} < \dots < 2^{-m} \langle \sigma \rangle_{P_0}$ by the construction of the principal cubes, and hence

$$\begin{aligned} \left[\sum_{P \in \mathcal{P}^a} \langle \sigma \rangle_P 1_{\{|S_{\mathcal{K}_b^a(P)}| > n2^{-b} \langle \sigma \rangle_P\}}(x) \right]^2 &= \left[\sum_{m=0}^{\infty} \langle \sigma \rangle_{P_m} \right]^2 \\ &\leq \left[\sum_{m=0}^{\infty} 2^{-m} \langle \sigma \rangle_{P_0} \right]^2 = 4 \langle \sigma \rangle_{P_0}^2 \\ &\leq 4 \sum_{P \in \mathcal{P}^a} \langle \sigma \rangle_P^2 1_{\{|S_{\mathcal{K}_b^a(P)}(\sigma 1_Q)| > n2^{-b} \langle \sigma \rangle_P\}}(x) \end{aligned}$$

Hence

$$\begin{aligned} &\left\| \sum_{P \in \mathcal{P}^a} \langle \sigma \rangle_P 1_{\{|S_{\mathcal{K}_b^a(P)}(\sigma 1_Q)| > n2^{-b} \langle \sigma \rangle_P\}} \right\|_{L^2(w)} \\ &\leq \left(\int \left[4 \sum_{P \in \mathcal{P}^a} \langle \sigma \rangle_P^2 1_{\{|S_{\mathcal{K}_b^a(P)}(\sigma 1_Q)| > n2^{-b} \langle \sigma \rangle_P\}} \right] w \right)^{1/2} \\ &= 2 \left(\sum_{P \in \mathcal{P}^a} \langle \sigma \rangle_P^2 w(\{|S_{\mathcal{K}_b^a(P)}(\sigma 1_Q)| > n2^{-b} \langle \sigma \rangle_P\}) \right)^{1/2}, \end{aligned}$$

and it remains to obtain good estimates for the measure of the level sets

$$\{|S_{\mathcal{K}_b^a(P)}(\sigma 1_Q)| > n2^{-b} \langle \sigma \rangle_P\}.$$

5.B. Weak-type and John–Nirenberg-style estimates. We still need to estimate the sets above. Recall that $S_{\mathcal{K}_b^a(P)}$ is a subshift of S , which in particular has its scales separated so that $\log_2 \ell(K) \equiv k \pmod{\kappa+1}$ for all K for which A_K participating in $S_{\mathcal{K}_b^a(P)}$ is nonzero and $k \in \{0, 1, \dots, \kappa := \max\{i, j\}\}$ is fixed, S being of type (i, j) . The following estimate deals with such subshifts, which we simply denote by S .

5.1. Proposition. *Let S be a dyadic shift of type (i, j) with scales separated. Then*

$$|\{|Sf| > \lambda\}| \leq \frac{C}{\lambda} \|f\|_1, \quad \forall \lambda > 0,$$

where C depends only on the dimension.

Proof. The proof uses the classical Calderón–Zygmund decomposition:

$$f = g + b, \quad b := \sum_{L \in \mathcal{B}} b_L := \sum_{L \in \mathcal{B}} 1_B(f - \langle f \rangle_L),$$

where $L \in \mathcal{B}$ are the maximal dyadic cubes with $\langle |f| \rangle_L > \lambda$; hence $\langle |f| \rangle_L \leq 2^d \lambda$. As usual,

$$g = f - b = 1_{(\cup \mathcal{B})^c} f + \sum_{L \in \mathcal{B}} \langle f \rangle_L$$

satisfies $\|g\|_\infty \leq 2^d \lambda$ and $\|g\|_1 \leq \|f\|_1$, hence $\|g\|_2^2 \leq \|g\|_\infty \|g\|_1 \leq 2^d \lambda \|f\|_1$, and thus

$$|\{ |Sg| > \tfrac{1}{2} \lambda \}| \leq \frac{4}{\lambda^2} \|Sg\|_2^2 \leq \frac{4}{\lambda^2} \|g\|_2^2 \leq 4 \cdot 2^d \frac{1}{\lambda} \|f\|_1.$$

It remains to estimate $\{ |Sb| > \tfrac{1}{2} \lambda \}$. First observe that

$$Sb = \sum_{K \in \mathcal{D}} \sum_{L \in \mathcal{B}} A_K b_L = \sum_{L \in \mathcal{B}} \left(\sum_{K \subseteq L} A_K b_L + \sum_{K \supsetneq L} A_K b_L \right),$$

since $A_K b_L \neq 0$ only if $K \cap L \neq \emptyset$. Now

$$\begin{aligned} |\{ |Sb| > \tfrac{1}{2} \lambda \}| &\leq \left| \left\{ \left| \sum_{L \in \mathcal{B}} \sum_{K \subseteq L} A_K b_L \right| > 0 \right\} \right| + \left| \left\{ \left| \sum_{L \in \mathcal{B}} \sum_{K \supsetneq L} A_K b_L \right| > \tfrac{1}{2} \lambda \right\} \right| \\ &\leq \sum_{L \in \mathcal{B}} |L| + \frac{2}{\lambda} \left\| \sum_{L \in \mathcal{B}} \sum_{K \supsetneq L} A_K b_L \right\|_1 \\ &\leq \frac{1}{\lambda} \|f\|_1 + \frac{2}{\lambda} \sum_{L \in \mathcal{B}} \sum_{K \supsetneq L} \|A_K b_L\|_1, \end{aligned}$$

where we used the elementary properties of the Calderón–Zygmund decomposition to estimate the first term.

For the remaining double sum, we still need some observations. Recall that

$$A_K b_L = \sum_{\substack{I, J \subseteq K \\ \ell(I) = 2^{-i} \ell(K) \\ \ell(J) = 2^{-j} \ell(K)}} a_{IJK} h_I \langle h_J, b_L \rangle.$$

Now, if $\ell(K) > 2^\kappa \ell(L) \geq 2^j \ell(L)$, then $\ell(J) > \ell(L)$, and hence h_J is constant on L . But the integral of b_L vanishes, hence $\langle h_J, b_L \rangle = 0$ for all relevant J , and thus $A_K b_L = 0$ whenever $\ell(K) > 2^\kappa \ell(L)$.

Thus, in the inner sum, the only possible nonzero terms are $A_K b_L$ for $K = L^{(m)}$ for $m = 1, \dots, \kappa$. By the separation of scales, at most one of these terms is nonzero, and we write \tilde{L} for the corresponding unique K . So in fact

$$\frac{2}{\lambda} \sum_{L \in \mathcal{B}} \sum_{K \supsetneq L} \|A_K b_L\|_1 = \frac{2}{\lambda} \sum_{L \in \mathcal{B}} \|A_{\tilde{L}} b_L\|_1 \leq \frac{2}{\lambda} \sum_{L \in \mathcal{B}} \|b_L\|_1 \leq \frac{2}{\lambda} \cdot 2 \|f\|_1 = \frac{4}{\lambda} \|f\|_1$$

by using the normalized boundedness of the averaging operators $A_{\tilde{L}}$ on $L^1(\mathbb{R}^d)$, and an elementary estimate for the bad part of the Calderón–Zygmund decomposition.

Altogether, we obtain the claim with $C = 4 \cdot 2^d + 5$. \square

For the special subshifts $S_{\mathcal{K}_b^a(P)}$, we can improve the weak-type $(1, 1)$ estimate to an exponential decay:

5.2. Proposition. *Let $S_{\mathcal{K}_b^a(P)}$ be the subshift of S as constructed earlier. Then the following estimate holds when ν is either the Lebesgue measure or w :*

$$\nu \left(\left\{ |S_{\mathcal{K}_b^a(P)}(\sigma 1_Q)| > C 2^{-b} \langle \sigma \rangle_P \cdot t \right\} \right) \lesssim C 2^{-t} \nu(P), \quad t \geq 0,$$

where C is a constant.

Proof. Let $\lambda := C2^{-b}\langle\sigma\rangle_P$, where C is a large constant, and $n \in \mathbb{Z}_+$. Let $x \in \mathbb{R}^d$ be a point where

$$(5.3) \quad |S_{\mathcal{K}_b^a(P)}(\sigma 1_Q)(x)| > n\lambda.$$

Then for all small enough $L \in \mathcal{K}_b^a(P)$ with $L \ni x$, there holds

$$\left| \sum_{\substack{K \in \mathcal{K}_b^a(P) \\ K \supseteq L}} A_K(\sigma 1_Q)(x) \right| > n\lambda.$$

Since $\sum_{\substack{K \in \mathcal{K}_b^a(P) \\ K \supseteq L}} A_K(\sigma 1_Q)$ is constant on L (thanks to separation of scales), and

$$(5.4) \quad \|A_L(\sigma 1_Q)\|_\infty \lesssim \frac{\sigma(L)}{|L|} \leq 2^{1-b} \frac{\sigma(P)}{|P|},$$

it follows that

$$(5.5) \quad \left| \sum_{\substack{K \in \mathcal{K}_b^a(P) \\ K \supseteq L}} A_K(\sigma 1_Q) \right| > (n - \frac{2}{3})\lambda \quad \text{on } L.$$

Let $\mathcal{L} \subseteq \mathcal{K}_b^a(P)$ be the collection of maximal cubes with the above property. Thus all $L \in \mathcal{L}$ are disjoint, and all x with (5.3) belong to some L . By maximality of L , the minimal $L^* \in \mathcal{K}_b^a(S)$ with $L^* \supsetneq L$ satisfies

$$\left| \sum_{\substack{K \in \mathcal{K}_b^a(P) \\ K \supseteq L^*}} A_K(\sigma 1_Q) \right| \leq (n - \frac{2}{3})\lambda \quad \text{on } L^*.$$

By an estimate similar to (5.4), with L^* in place of L , it follows that

$$\left| \sum_{\substack{K \in \mathcal{K}_b^a(P) \\ K \supseteq L}} A_K(\sigma 1_Q) \right| \leq (n - \frac{1}{3})\lambda \quad \text{on } L.$$

Thus, if x satisfies (5.3) and $x \in L \in \mathcal{L}$, then necessarily

$$|S_{\{K \in \mathcal{K}_b^a(P); K \subseteq L\}}(\sigma 1_Q)(x)| = \left| \sum_{\substack{K \in \mathcal{K}_b^a(P) \\ K \subseteq L}} A_K(\sigma 1_Q)(x) \right| > \frac{1}{3}\lambda.$$

Using the weak-type L^1 estimate to the shift $S_{\{K \in \mathcal{K}_b^a(P); K \subseteq L\}}$ of type (i, j) with scales separated, noting that $A_K(\sigma 1_Q) = A_K(\sigma 1_L)$ for $K \subseteq L$, it follows that

$$\begin{aligned} \left| \left\{ \left| \sum_{\substack{K \in \mathcal{K}_b^a(P) \\ K \subseteq L}} A_K(\sigma 1_Q)(x) \right| > \frac{1}{3}\lambda \right\} \right| &\leq \frac{C}{\lambda} \sigma(L) \\ &\leq \frac{C}{\lambda} 2^{1-b} \frac{\sigma(S \cap Q)}{|S|} |L| \leq \frac{1}{3} |L|, \end{aligned}$$

provided that the constant in the definition of λ was chosen large enough. Recalling (5.5), there holds

$$\begin{aligned} \left| \sum_{K \in \mathcal{K}_b^a(P)} A_K(\sigma 1_Q) \right| &\geq \left| \sum_{\substack{K \in \mathcal{K}_b^a(P) \\ K \supseteq L}} A_K(\sigma 1_Q) \right| - \left| \sum_{\substack{K \in \mathcal{K}_b^a(P) \\ K \subseteq L}} A_K(\sigma 1_Q) \right| \\ &> (n - \frac{2}{3})\lambda - \frac{1}{3}\lambda = (n - 1)\lambda \quad \text{on } \tilde{L} \subset L \text{ with } |\tilde{L}| \geq \frac{2}{3}|L|. \end{aligned}$$

Thus

$$\begin{aligned}
|\{|S_{\mathcal{K}_b^a(P)}(\sigma 1_Q)| > n\lambda\}| &\leq \sum_{L \in \mathcal{L}} |L \cap \{|S_{\mathcal{K}_b^a(P)}(\sigma 1_Q)| > n\lambda\}| \\
&\leq \sum_{L \in \mathcal{L}} |\{|S_{\{K \in \mathcal{K}_b^a(P) : K \subseteq L\}}(\sigma 1_Q)| > \tfrac{1}{3}\lambda\}| \\
&\leq \sum_{L \in \mathcal{L}} \tfrac{1}{3}|L| \leq \sum_{L \in \mathcal{L}} \tfrac{1}{3} \cdot \tfrac{3}{2}|\tilde{L}| \\
&\leq \tfrac{1}{2} \sum_{L \in \mathcal{L}} |L \cap \{|S_{\mathcal{K}_b^a(P)}(\sigma 1_Q)| > (n-1)\lambda\}| \\
&\leq \tfrac{1}{2} |\{|S_{\mathcal{K}_b^a(P)}(\sigma 1_Q)| > (n-1)\lambda\}|.
\end{aligned}$$

By induction it follows that

$$\begin{aligned}
|\{|S_{\mathcal{K}_b^a(P)}(\sigma 1_Q)| > n\lambda\}| &\leq 2^{-n} |\{|S_{\mathcal{K}_b^a(P)}(\sigma 1_Q)| > 0\}| \\
&\leq 2^{-n} \sum_{M \in \mathcal{M}} |M| \leq 2^{-n} |P|,
\end{aligned}$$

where \mathcal{M} is the collection of maximal cubes in $\mathcal{K}_b^a(S)$.

To deduce the corresponding estimate for the w measure, selected intermediate steps of the above computation, as well as the definition of $\mathcal{K}_b^a(P)$, will be exploited. Recall that $K \in \mathcal{K}^a$ means that $2^{a-1} < \langle w \rangle_K \langle \sigma \rangle_K \leq 2^a$, while $K \in \mathcal{K}_b^a(P)$ means that in addition $2^{-b} < \langle \sigma \rangle_K / \langle \sigma \rangle_P \leq 2^{1-b}$. Put together, this says that

$$2^{a+b-2} \langle \sigma \rangle_P < \frac{w(K)}{|K|} < 2^{a+b} \langle \sigma \rangle_P \quad \forall K \in \mathcal{K}_b^a(P).$$

Hence, using the collections $\mathcal{L}, \mathcal{M} \subseteq \mathcal{K}_b^a(P)$ as above,

$$\begin{aligned}
w(\{| \text{III}_{\mathcal{K}_b^a(P)}(\sigma 1_Q) | > n\lambda\}) &\leq \sum_{L \in \mathcal{L}} w(L) \leq \sum_{L \in \mathcal{L}} 2^{a+b} \langle \sigma \rangle_P |L| \\
&\leq 2^{a+b} \langle \sigma \rangle_P |\{| \text{III}_{\mathcal{K}_b^a(P)}(\sigma 1_Q) | > (n-1)\lambda\}| \\
&\leq 2^{a+b} \langle \sigma \rangle_P \cdot 2^{-n} \sum_{M \in \mathcal{M}} |M| \\
&\leq 4 \cdot 2^{-n} \sum_{M \in \mathcal{M}} w(M) \leq 4 \cdot 2^{-n} w(S). \quad \square
\end{aligned}$$

5.C. Conclusion of the estimation of the testing conditions. Recall that

$$\begin{aligned}
&\left\| \sum_{K \subseteq Q} A_K(\sigma 1_Q) \right\|_{L^2(w)} \\
&\leq \sum_{k=0}^{\kappa} \sum_a \sum_{b=0}^{\infty} 2^{-b} \sum_{n=0}^{\infty} (1+n) \left\| \sum_{P \in \mathcal{P}^a} \langle \sigma \rangle_P 1_{\{|S_{\mathcal{K}_b^a(P)}(\sigma 1_Q)| > n2^{-b} \langle \sigma \rangle_P\}} \right\|_{L^2(w)}
\end{aligned}$$

and

$$\begin{aligned}
& \left\| \sum_{P \in \mathcal{P}^a} \langle \sigma \rangle_P 1_{\{|S_{\mathcal{H}_b^a(P)}(\sigma 1_Q)| > n 2^{-b} \langle \sigma \rangle_P\}} \right\|_{L^2(w)} \\
& \leq 2 \left(\sum_{P \in \mathcal{P}^a} \langle \sigma \rangle_P^2 w(\{|S_{\mathcal{H}_b^a(P)}(\sigma 1_Q)| > n 2^{-b} \langle \sigma \rangle_P\}) \right)^{1/2} \\
& \leq C \left(\sum_{P \in \mathcal{P}^a} \langle \sigma \rangle_P^2 2^{-n/C} w(P) \right)^{1/2} \\
& = C 2^{-cn} \left(\sum_{P \in \mathcal{P}^a} \frac{\sigma(P) w(P)}{|P|^2} \sigma(P) \right)^{1/2} \\
& \leq C 2^{-cn} \left(2^a \sum_{P \in \mathcal{P}^a} \sigma(P) \right)^{1/2},
\end{aligned}$$

recalling the freezing of the A_2 characteristic between 2^{a-1} and 2^a for cubes in $\mathcal{H}^a \supseteq \mathcal{P}^a$.

For the summation over the principal cubes, we observe that

$$\sum_{P \in \mathcal{P}^a} \sigma(P) = \sum_{P \in \mathcal{P}^a} \langle \sigma \rangle_P |P| = \int_Q \sum_{P \in \mathcal{P}^a} \langle \sigma \rangle_P 1_P(x) dx.$$

At any given x , if $P_0 \subsetneq P_1 \subsetneq \dots \subseteq Q$ are the principal cubes containing it, we have

$$\sum_{P \in \mathcal{P}^a} \langle \sigma \rangle_P 1_P(x) = \sum_{m=0}^{\infty} \langle \sigma \rangle_{P_m} \leq \sum_{m=0}^{\infty} 2^{-m} \langle \sigma \rangle_{P_0} = 2 \langle \sigma \rangle_{P_0} \leq 2M(\sigma 1_Q)(x),$$

where M is the dyadic maximal operator. Hence

$$\sum_{P \in \mathcal{P}^a} \sigma(P) \leq 2 \int_Q M(\sigma 1_Q) dx \leq 2[\sigma]_{A_\infty} \sigma(Q)$$

by the definition of the A_∞ characteristic

$$[\sigma]_{A_\infty} := \sup_Q \frac{1}{\sigma(Q)} \int_Q M(\sigma 1_Q) dx.$$

Substituting back, we have

$$\begin{aligned}
& \left\| \sum_{K \subseteq Q} A_K(\sigma 1_Q) \right\|_{L^2(w)} \\
& \leq \sum_{k=0}^{\kappa} \sum_a \sum_{b=0}^{\infty} 2^{-b} \sum_{n=0}^{\infty} (1+n) \left\| \sum_{P \in \mathcal{P}^a} \langle \sigma \rangle_P 1_{\{|S_{\mathcal{H}_b^a(P)}(\sigma 1_Q)| > n 2^{-b} \langle \sigma \rangle_P\}} \right\|_{L^2(w)} \\
& \leq \sum_{k=0}^{\kappa} \sum_a \sum_{b=0}^{\infty} 2^{-b} \sum_{n=0}^{\infty} (1+n) \cdot C 2^{-cn} \left(2^a \sum_{P \in \mathcal{P}^a} \sigma(P) \right)^{1/2} \\
& \leq \sum_{k=0}^{\kappa} \sum_a \sum_{b=0}^{\infty} 2^{-b} \sum_{n=0}^{\infty} (1+n) \cdot C 2^{-cn} \left(2^a [\sigma]_{A_\infty} \right)^{1/2} \\
& = C \cdot [\sigma]_{A_\infty}^{1/2} \sum_{k=0}^{\kappa} \left(\sum_{a \leq \lceil \log_2[w, \sigma]_{A_2} \rceil} 2^{a/2} \right) \left(\sum_{b=0}^{\infty} 2^{-b} \right) \left(\sum_{n=0}^{\infty} (1+n) \cdot 2^{-cn} \right) \\
& \leq C \cdot [\sigma]_{A_\infty}^{1/2} \cdot (1+\kappa) \cdot [w, \sigma]_{A_2}^{1/2},
\end{aligned}$$

and thus the testing constant \mathfrak{S} is estimated by

$$\mathfrak{S} \leq C \cdot (1 + \kappa) \cdot [w, \sigma]_{A_2}^{1/2} \cdot [\sigma]_{A_\infty}^{1/2}.$$

By symmetry, exchanging the roles of w and σ , we also have the analogous result for \mathfrak{S}^* , and so we have completed the proof of the following:

5.6. Theorem. *Let $\sigma, w \in A_\infty$ be functions which satisfy the joint A_2 condition*

$$[w, \sigma]_{A_2} := \sup_Q \frac{w(Q)\sigma(Q)}{|Q|^2} < \infty.$$

Then the testing constants \mathfrak{S} and \mathfrak{S}^ associated with a dyadic shift S of type (i, j) satisfy the following bounds, where $\kappa := \max\{i, j\}$:*

$$\begin{aligned} \mathfrak{S} &\leq C \cdot (1 + \kappa) \cdot [w, \sigma]_{A_2}^{1/2} \cdot [\sigma]_{A_\infty}^{1/2}, \\ \mathfrak{S}^* &\leq C \cdot (1 + \kappa) \cdot [w, \sigma]_{A_2}^{1/2} \cdot [w]_{A_\infty}^{1/2}. \end{aligned}$$

6. CONCLUSIONS

In this section we simply collect the fruits of the hard work done above. A combination of Theorem 4.2 and 5.6 gives the following two-weight inequality, whose qualitative version was pointed out by Lacey, Petermichl and Reguera [11]:

6.1. Theorem. *Let $\sigma, w \in A_\infty$ be functions which satisfy the joint A_2 condition*

$$[w, \sigma]_{A_2} := \sup_Q \frac{w(Q)\sigma(Q)}{|Q|^2} < \infty.$$

Then a dyadic shift S of type (i, j) satisfies $S(\sigma \cdot) : L^2(\sigma) \rightarrow L^2(w)$, and more precisely

$$\|S(\sigma \cdot)\|_{L^2(\sigma) \rightarrow L^2(w)} \lesssim (1 + \kappa)^2 [w, \sigma]_{A_2}^{1/2} ([w]_{A_\infty}^{1/2} + [\sigma]_{A_\infty}^{1/2}),$$

where $\kappa = \max\{i, j\}$.

The quantitative bound as stated, including the polynomial dependence on κ , allows to sum up these estimates in the Dyadic Representation Theorem to deduce:

6.2. Theorem. *Let $\sigma, w \in A_\infty$ be functions which satisfy the joint A_2 condition. Then any L^2 bounded Calderón–Zygmund operator T whose kernel K has Hölder type modulus of continuity $\psi(t) = t^\alpha$, $\alpha \in (0, 1)$, satisfies*

$$\|T(\sigma \cdot)\|_{L^2(\sigma) \rightarrow L^2(w)} \lesssim (\|T\|_{L^2 \rightarrow L^2} + \|K\|_{CZ_\alpha}) [w, \sigma]_{A_2}^{1/2} ([w]_{A_\infty}^{1/2} + [\sigma]_{A_\infty}^{1/2}).$$

Recalling the dual weight trick and specializing to the one-weight situation with $\sigma = w^{-1}$, this in turn gives:

6.3. Theorem. *Let $w \in A_2$. Then any L^2 bounded Calderón–Zygmund operator T whose kernel K has Hölder type modulus of continuity $\psi(t) = t^\alpha$, $\alpha \in (0, 1)$, satisfies*

$$\begin{aligned} \|T\|_{L^2(w) \rightarrow L^2(w)} &\lesssim (\|T\|_{L^2 \rightarrow L^2} + \|K\|_{CZ_\alpha}) [w]_{A_2}^{1/2} ([w]_{A_\infty}^{1/2} + [w^{-1}]_{A_\infty}^{1/2}) \\ &\lesssim (\|T\|_{L^2 \rightarrow L^2} + \|K\|_{CZ_\alpha}) [w]_{A_2}. \end{aligned}$$

The second displayed line is the original A_2 theorem [6], and it follows from the first line by $[w]_{A_\infty} \lesssim [w]_{A_2}$ and $[w^{-1}]_{A_\infty} \lesssim [w^{-1}]_{A_2} = [w]_{A_2}$. Its strengthening on the first line was first observed in my joint work with C. Pérez [9]. Note that, compared to the introductory statement in Theorem 1.1, the dependence on the operator T has been made more explicit. (The implied constants in the notation “ \lesssim ” only depend on the dimension and the Hölder exponent α .) This dependence on $\|T\|_{L^2 \rightarrow L^2}$ and $\|K\|_{CZ_\alpha}$ is implicit in the original proof, but has not been spelled out before.

7. FURTHER RESULTS AND REMARKS

This final section briefly collects some related developments, which were not covered in the actual lectures.

The A_2 theorem implies a corresponding A_p theorem for all $p \in (1, \infty)$. This follows from the sharp weighted extrapolation theorem of Dragičević, Grafakos, Pereyra, and Petermichl [3], which was known well before the proof of the full A_2 theorem:

7.1. Theorem. *If an operator T satisfies*

$$\|T\|_{L^2(w) \rightarrow L^2(w)} \leq C_T [w]_{A_2}^\tau$$

for all $w \in A_2$, then it satisfies

$$\|T\|_{L^p(w) \rightarrow L^p(w)} \leq c_p C_T [w]_{A_p}^{\tau \max\{1, 1/(p-1)\}}$$

for all $p \in (1, \infty)$ and $w \in A_p$.

7.2. Corollary. *Let $p \in (1, \infty)$ and $w \in A_p$. Then any L^2 bounded Calderón–Zygmund operator T whose kernel K has Hölder type modulus of continuity $\psi(t) = t^\alpha$, $\alpha \in (0, 1)$, satisfies*

$$\|T\|_{L^p(w) \rightarrow L^p(w)} \lesssim (\|T\|_{L^2 \rightarrow L^2} + \|K\|_{CZ_\alpha}) [w]_{A_p}^{\max\{1, 1/(p-1)\}}.$$

It is also possible to apply a version of the extrapolation argument to the mixed A_2/A_∞ bounds [9], but this did not give the optimal results for $p \neq 2$. However, by setting up a different argument directly in $L^p(w)$, the following bounds were obtained in my collaboration with M. Lacey [7]:

7.3. Theorem. *Let $p \in (1, \infty)$ and $w \in A_p$. Then any L^2 bounded Calderón–Zygmund operator T whose kernel K has Hölder type modulus of continuity $\psi(t) = t^\alpha$, $\alpha \in (0, 1)$, satisfies*

$$\|T\|_{L^p(w) \rightarrow L^p(w)} \lesssim (\|T\|_{L^2 \rightarrow L^2} + \|K\|_{CZ_\alpha}) [w]_{A_p}^{1/p} ([w]_{A_\infty}^{1/p'} + [w^{1-p'}]_{A_\infty}^{1/p}).$$

For weak-type bounds, which were investigated by Lacey, Martikainen, Orponen, Reguera, Sawyer, Uriarte-Tuero, and myself [8], we need only ‘half’ of the strong-type upper bound:

7.4. Theorem. *Let $p \in (1, \infty)$ and $w \in A_p$. Then any L^2 bounded Calderón–Zygmund operator T whose kernel K has Hölder type modulus of continuity $\psi(t) = t^\alpha$, $\alpha \in (0, 1)$, satisfies*

$$\begin{aligned} \|T\|_{L^p(w) \rightarrow L^{p,\infty}(w)} &\lesssim (\|T\|_{L^2 \rightarrow L^2} + \|K\|_{CZ_\alpha}) [w]_{A_p}^{1/p} [w]_{A_\infty}^{1/p'} \\ &\lesssim (\|T\|_{L^2 \rightarrow L^2} + \|K\|_{CZ_\alpha}) [w]_{A_p}. \end{aligned}$$

All these results remain valid for the *maximal truncations*

$$T_{\sharp}f(x) := \sup_{\varepsilon > 0} |T_{\varepsilon}f(x)|, \quad T_{\varepsilon}f(x) := \int_{|x-y|>\varepsilon} K(x,y)f(y) \, dy,$$

which have been addressed in [7, 8]. In [8] it was also shown that the sharp weighted bounds for dyadic shifts can be made linear (instead of quadratic) in κ , a result recovered by a different (Bellman function) method by Treil [20]. Earlier polynomial-in- κ Bellman function estimates for the shifts were due to Nazarov and Volberg [17]. An extension of the A_2 theorem to metric spaces with a doubling measure (spaces of homogeneous type) is due to Nazarov, Reznikov, and Volberg [14].

It seems that the strategy developed in [7] is the most efficient one for dealing with both the maximal truncations and the mixed A_p/A_{∞} bounds. It consists of the following steps:

- Reduction of Calderón–Zygmund operators to dyadic shift operators by the Representation Theorem, just like here.
- Reduction of the dyadic shifts to *positive dyadic shifts* by an ingenious formula of A. Lerner [13], which provides precise and useful information of a function in terms of its ‘local oscillations’: the local oscillations of Sf (and even $S_{\sharp}f$) can be estimated in terms of f with the help of the weak-type $(1,1)$ inequality for S (and S_{\sharp}).
- Reduction of the estimates for positive shifts to testing conditions.
- Verification of the testing conditions for the positive shifts.

Notice that these steps are essentially the same as the ones followed in these lectures, except for the additional second step. After this further reduction, the last two steps are even slightly easier, since they are only needed for positive operators. In this setting, the orthogonality arguments, which were decisive for the present reduction to testing conditions, are replaced by positive-kernel estimates, for which appropriate theory valid in all L^p , not just $p = 2$, has been developed by Lacey, Sawyer, and Uriarte-Tuero [12].

7.A. The Ahlfors–Beurling operator and its powers. One of the key original motivations to study the A_2 theorem was a conjecture of Astala–Iwaniec–Saksman [1] concerning the special case where T is the Ahlfors–Beurling operator

$$Bf(z) := -\frac{1}{\pi} \text{p. v.} \int_{\mathbb{C}} \frac{1}{\zeta^2} f(z - \zeta) \, dA(\zeta),$$

and A is the area measure (two-dimensional Lebesgue measure) on $\mathbb{C} \simeq \mathbb{R}^2$. This was the first Calderón–Zygmund operator for which the A_2 theorem was proven; it was achieved by Petermichl and Volberg [19], confirming the mentioned conjecture of Astala, Iwaniec, and Saksman [1]. Another proof of the A_2 theorem for this specific operator is due to Dragičević and Volberg [4].

The powers B^n of B have also been studied, and then it is of interest to understand the growth of the norms as a function of n . Shortly before the proof of the full A_2 theorem, by methods specific to the Ahlfors–Beurling operator, O. Dragičević [2] was able to prove the cubic growth

$$\|B^n\|_{L^p(w) \rightarrow L^p(w)} \lesssim |n|^3 [w]_{A_p}^{\max\{1, 1/(p-1)\}}, \quad n \in \mathbb{Z} \setminus \{0\}.$$

Now, let us see what the general A_2 theorem gives for these specific powers. It is known that B^n is the convolution operator with the kernel

$$K_n(z) = (-1)^n \frac{|n|}{\pi} \left(\frac{\bar{z}}{z} \right)^n |z|^{-2},$$

and it is elementary to check that this satisfies $\|K_n\|_{CZ_\alpha} \lesssim |n|^{1+\alpha}$ for any $\alpha \in (0, 1)$. Moreover, since B is an isometry on $L^2(\mathbb{C})$, we have $\|B^n\|_{L^2 \rightarrow L^2} = 1$. From Corollary 7.2 we deduce:

7.5. Corollary. *The powers B^n of the Ahlfors–Beurling operator satisfy*

$$\|B^n\|_{L^p(w) \rightarrow L^p(w)} \lesssim |n|^{1+\alpha} [w]_{A_p}^{\max\{1, 1/(p-1)\}}, \quad \alpha > 0,$$

where the implied constant depends on p and α .

Thus the cubic bound improves arbitrarily close to a linear one. It seems plausible that even the linear growth ($\alpha = 0$) should be true, but this would require additional insight, probably specific to the operator B . This final corollary is new; it arose from my discussions with O. Dragičević in July 2011. In fact, this application motivated the formulation of the Dyadic Representation Theorem and the A_2 theorem with explicit dependence on $\|T\|_{L^2 \rightarrow L^2} + \|K\|_{CZ_\alpha}$, which might also turn out useful in other applications to families of Calderón–Zygmund operators.

REFERENCES

- [1] Kari Astala, Tadeusz Iwaniec, and Eero Saksman. Beltrami operators in the plane. *Duke Math. J.*, 107(1):27–56, 2001.
- [2] Oliver Dragičević. Weighted estimates for powers and the Ahlfors–Beurling operator. *Proc. Amer. Math. Soc.*, 139(6):2113–2120, 2011.
- [3] Oliver Dragičević, Loukas Grafakos, María Cristina Pereyra, and Stefanie Petermichl. Extrapolation and sharp norm estimates for classical operators on weighted Lebesgue spaces. *Publ. Mat.*, 49(1):73–91, 2005.
- [4] Oliver Dragičević and Alexander Volberg. Sharp estimate of the Ahlfors–Beurling operator via averaging martingale transforms. *Michigan Math. J.*, 51(2):415–435, 2003.
- [5] Tuomas Hytönen. The vector-valued non-homogeneous Tb theorem. *Preprint*, arXiv:0809.3097, 2008.
- [6] Tuomas Hytönen. The sharp weighted bound for general Calderón–Zygmund operators. *Ann. of Math. (2)*, to appear in 2012. Preprint, arXiv:1007.4330, 2010.
- [7] Tuomas Hytönen and Michael T. Lacey. The A_p – A_∞ inequality for general Calderón–Zygmund operators. *Preprint*, arXiv:1106.4797, 2011.
- [8] Tuomas Hytönen, Michael T. Lacey, Henri Martikainen, Tuomas Orponen, Maria Carmen Reguera, Eric T. Sawyer, and Ignacio Uriarte-Tuero. Weak and strong type estimates for maximal truncations of Calderón–Zygmund operators on A_p weighted spaces. *Preprint*, arXiv:1103.5229, 2011.
- [9] Tuomas Hytönen and Carlos Pérez. Sharp weighted bounds involving A_∞ . *Preprint*, arXiv:1103.5562, 2011.
- [10] Tuomas Hytönen, Carlos Pérez, Sergei Treil, and Alexander Volberg. Sharp weighted estimates for dyadic shifts and the A_2 conjecture. *Preprint*, arXiv:1010.0755, 2010.
- [11] Michael T. Lacey, Stefanie Petermichl, and Maria Carmen Reguera. Sharp A_2 inequality for Haar shift operators. *Math. Ann.*, 348(1):127–141, 2010.
- [12] Michael T. Lacey, Eric T. Sawyer, and Ignacio Uriarte-Tuero. Two weight inequalities for discrete positive operators. Preprint, arXiv:0911.3437 (2009).
- [13] Andrei K. Lerner. A pointwise estimate for the local sharp maximal function with applications to singular integrals. *Bull. Lond. Math. Soc.*, 42(5):843–856, 2010.
- [14] Fedor Nazarov, Alexander Reznikov, and Alexander Volberg. The proof of A_2 conjecture in a geometrically doubling metric space. *Preprint*, arXiv:1106.1342, 2011.

- [15] Fedor Nazarov, Sergei Treil, and Alexander Volberg. The Tb -theorem on non-homogeneous spaces. *Acta Math.*, 190(2):151–239, 2003.
- [16] Fedor Nazarov, Sergei Treil, and Alexander Volberg. Two weight inequalities for individual Haar multipliers and other well localized operators. *Math. Res. Lett.*, 15(3):583–597, 2008.
- [17] Fedor Nazarov and Alexander Volberg. A simple sharp weighted estimate of the dyadic shifts on metric spaces with geometric doubling. *Preprint*, arXiv:1104.4893, 2011.
- [18] Carlos Pérez, Sergei Treil, and Alexander Volberg. On A_2 conjecture and corona decomposition of weights. *Preprint*, arXiv:1006.2630, 2010.
- [19] Stefanie Petermichl and Alexander Volberg. Heating of the Ahlfors-Beurling operator: weakly quasiregular maps on the plane are quasiregular. *Duke Math. J.*, 112(2):281–305, 2002.
- [20] Sergei Treil. Sharp A_2 estimates of Haar shifts via Bellman function. *Preprint*, arXiv:1105.2252, 2011.

DEPARTMENT OF MATHEMATICS AND STATISTICS, P.O.B. 68 (GUSTAF HÄLLSTRÖMIN KATU 2B),
 FI-00014 UNIVERSITY OF HELSINKI, FINLAND
E-mail address: `tuomas.hytonen@helsinki.fi`